



Ibraimo Mamudo Osmane Alegy

**Previsão da Vida Útil Restante para a Manutenção Preditiva das
Baterias de Veículos Elétricos:
Um Contributo com Conjuntos de Dados *Benchmark***

Coimbra, outubro de 2025



Ibraimo Mamudo Osmane Alegy

**Previsão da Vida Útil Restante para a Manutenção
Preditiva das Baterias de Veículos Elétricos:
Um Contributo com Conjuntos de Dados *Benchmark***

Dissertação submetida ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de **Mestre em Análise de Dados e Sistemas de Apoio à Decisão**, realizada sob a orientação da Professora Doutora Joana Jorge de Queiroz Leite.

Coimbra, outubro de 2025

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

TERMO DE RESPONSABILIDADE

Declaro ser o autor desta dissertação, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau académico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação da presente dissertação.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

AGRADECIMENTOS

Agradeço à Professora Doutora Joana Leite pela paciência, pelo acompanhamento e pela orientação ao longo do desenvolvimento desta dissertação.

Agradeço também aos meus pais, que, mesmo à distância, estiveram sempre presentes nos momentos difíceis e me motivaram a não desistir, acreditando em mim e nas minhas capacidades.

Agradeço, ainda, aos meus amigos Marta Simões, Kevin Henrique, Ricardo Neto, Carlos Manjate e Daniela Ferreira pela disponibilidade constante para ajudar e por me motivarem a perseguir os meus objetivos.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

RESUMO

A previsão do RUL das baterias é um fator fundamental para a manutenção preditiva dos veículos elétricos contribuindo para a maximização do uso das baterias, minimizando assim a substituição antecipada de baterias que ainda não atingiram o limite de vida. O objetivo da presente dissertação passa pela realização da previsão do RUL, com recurso ao conjunto *benchmark*, treinando dados de degradação de três baterias do tipo iões de lítio para prever a vida útil restante de uma bateria considerada como desconhecida. Para o estudo, foram selecionados três modelos de aprendizagem computacional, nomeadamente o a regressão linear, o *random forest* e o XGBoost. Para avaliar o desempenho dos modelos, recorreu-se a três métricas de avaliação. Os resultados mostram que os modelos baseados em árvore, tendem a superar a regressão linear e que a normalização da vida útil restante melhora a generalização para a bateria alvo, sendo que o modelo que teve melhor desempenho de previsão foi o XGBoost.

Palavras-chave: Baterias de veículos elétricos, manutenção preditiva, aprendizagem computacional, vida útil restante, *BatteryAgingARC-FY08Q4*

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

ABSTRACT

Battery RUL prediction is a key factor in the predictive maintenance of electric vehicles, contributing to maximizing battery use and minimizing the early replacement of batteries that have not yet reached the end of their life cycle. The objective of this dissertation is to predict RUL using the benchmark dataset, training degradation data from three lithium-ion batteries to predict the remaining useful life of a battery considered unknown. Three machine learning models were selected for the study, namely linear regression, random forest, and XGBoost. Three evaluation metrics were selected to assess the performance of the models. The results show that tree-based models tend to outperform linear regression and that normalizing the remaining useful life improves generalization for the target battery, with XGBoost being the model with the best prediction performance.

Keywords: Electric vehicles batteries, predictive maintenance, machine learning, remaining useful life, *BatteryAgingARC-FY08Q4*

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

ÍNDICE GERAL

INTRODUÇÃO.....	1
1 REVISÃO DA LITERATURA.....	5
1.1 Manutenção preditiva.....	5
1.1.1 Benefícios da manutenção preditiva.....	7
1.1.2 Implementação da estratégia da PdM.....	8
1.2 RUL das baterias.....	9
1.3 As baterias nos veículos elétricos	10
1.3.1 Benefícios de baterias de iões de lítio	11
1.3.2 Características das baterias de iões de lítio	12
1.3.3 Degradação do ciclo das baterias.....	14
1.4 Abordagem orientada por dados	15
1.4.1 Abordagens baseadas em ML.....	16
1.4.2 Abordagens baseadas em aprendizagem profunda.....	19
1.4.3 Casos de aplicação para a previsão do SoH	20
1.4.4 Casos de aplicação para a previsão do RUL.....	22
1.5 Dados <i>benchmark</i>	25
2 ENTENDIMENTO DO PROBLEMA E DOS DADOS.....	28
2.1 Entendimento do problema.....	28
2.2 Entendimento dos dados	29
2.2.1 Estatísticas descritivas da bateria B0005.....	32
2.2.2 Estatísticas descritivas da bateria B0006.....	34

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

3	PREPARAÇÃO DOS DADOS	36
3.1	Limpeza dos dados.....	36
3.2	Extração de variáveis	38
3.3	Matriz de correlação	48
3.4	Seleção de variáveis.....	49
4	MODELAÇÃO.....	51
4.1	Modelos	51
4.1.1	Modelo <i>baseline</i>	51
4.1.2	RF	52
4.1.3	XGBoost.....	52
4.2	Hiperparâmetros.....	54
4.2.1	Hiperparâmetros do algoritmo RF	55
4.2.2	Hiperparâmetros do algoritmo XGBoost.....	56
4.3	Iterações	57
4.3.1	Iteração 1	57
4.3.2	Iteração 2	57
4.3.3	Iteração 3	58
5	RESULTADOS	59
5.1	Métricas de avaliação.....	59
5.2	Iteração 1.....	60
5.2.1	Modelo <i>baseline</i>	60
5.2.2	RF	61
5.2.3	XGBoost.....	62

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

5.2.4	Comparação dos resultados das métricas	62
5.3	Iteração 2.....	63
5.3.1	Otimização dos hiperparâmetros	63
5.3.2	Treinar os algoritmos RF Otimizado	66
5.3.3	Treinar o algoritmo XGBoost Otimizado.....	68
5.3.4	Comparação entre os algoritmos antes e depois da otimização.....	69
5.4	Iteração 3.....	69
5.4.1	Modelo <i>baseline</i>	69
5.4.2	RF	70
5.4.3	XGBoost.....	71
5.4.4	Comparação dos resultados das métricas dos modelos	72
CONCLUSÃO.....		73
REFERÊNCIAS BIBLIOGRÁFICAS		76
APÊNDICES		83
APÊNDICE 1. Estatísticas descritivas		84
APÊNDICE 2. Pipeline da modelação		86

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

ÍNDICE DE TABELAS

Tabela 1.1 Casos de aplicação de modelos de ML e/ou de DL para a previsão do SoH....	22
Tabela 1.2 Casos de aplicação de modelos de ML e/ou de DL para a previsão do RUL...	24
Tabela 2.1 Informação existente em cada um do tipo ciclo.	31
Tabela 2.2 Estatísticas descritivas bateria B0005.....	32
Tabela 2.3 Estatísticas descritivas bateria B0006.....	34
Tabela 3.1 Variáveis de carga e descarga.....	39
Tabela 3.2 Estatísticas descritivas das variáveis criadas	47
Tabela 4.1 Hiperparâmetros do algoritmo RF	56
Tabela 4.2 Hiperparâmetros do algoritmo XGBoost.....	57
Tabela 5.1 Comparação das métricas - Iteração 1	63
Tabela 5.2 GridSearchCV vs Optuna - RF	66
Tabela 5.3 GridSearchCV vs Optuna - XGBoost.....	66
Tabela 5.4 RF vs RF Otimizado	67
Tabela 5.5 XGBoost vs XGBoost Otimizado.....	68
Tabela 5.6 Comparação das métricas dos modelos - Iteração 3	72
Tabela 0.1 Estatísticas descritivas da bateria B0007	84
Tabela 0.2 Estatísticas descritivas da bateria B0018	85

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

ÍNDICE DE FIGURAS

Figura 1.1 Evolução das estratégias de manutenção ao longo do tempo	7
Figura 1.2 Célula de íons de lítio em descarregamento. Importa referir que o processo de carregamento ocorre no oposto.....	13
Figura 2.1 Degradação da capacidade das baterias	30
Figura 3.1 Ciclos inconsistentes de carga (12 e 13)	37
Figura 3.2 Ciclos inconsistentes de carga (32 e 33)	37
Figura 3.3 Ciclos inconsistentes de descarga (89 e 90).....	37
Figura 3.4 Ciclos inconsistentes de carga (46 e 47)	37
Figura 3.5 Ciclos inconsistentes de carga (57 e 58)	37
Figura 3.6 Total número de registos do ciclo 170 de carga	38
Figura 3.7 Voltagem medida ao longo de três ciclos de carga (a) e ao longo de três ciclos de descarga (b)	40
Figura 3.8 Corrente medida de três ciclos de carga (a) e três ciclos de descarga (b).....	41
Figura 3.9 Temperatura medida de três ciclos de carga (a) e três ciclos de descarga (b) ..	42
Figura 3.10 Corrente na fonte de carregamento de três ciclos de carga (a) e três ciclos de descarga (b)	43
Figura 3.11 Voltagem máxima na fonte de três ciclos de carga (a) e três ciclos de descarga (b)	44
Figura 3.12 Tempo total de cada ciclo de carga (a) e de descarga (b)	45
Figura 3.13 Matriz de correlação	48
Figura 5.1 RUL previsto vs RUL real - modelo <i>baseline</i>	61
Figura 5.2 RUL previsto vs RUL real - modelo RF	61
Figura 5.3 RUL previsto vs RUL real - modelo XGBoost.....	62

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Figura 5.4 Grelha de hiperparâmetros - RF	64
Figura 5.5 Grelha de hiperparâmetros - XGBoost.....	64
Figura 5.6 Intervalo para cada hiperparâmetro - RF	65
Figura 5.7 Intervalo para cada hiperparâmetro - XGBoost	65
Figura 5.8 RUL previsto vs RUL real - RF Otimizado	67
Figura 5.9 RUL previsto vs RUL real - XGBoost Otimizado.....	68
Figura 5.10 RUL previsto vs RUL real normalizado - modelo baseline.....	70
Figura 5.11 RUL previsto vs RUL real normalizado - algoritmo RF	71
Figura 5.12 RUL previsto vs RUL real normalizado - algoritmo XGBoost	71

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Lista de abreviaturas, acrónimos e siglas

Sigla	Português	Inglês
ANN	Rede Neuronal Artificial	<i>Artificial Neural Network</i>
BEV	Veículo Elétrico a Bateria	<i>Battery Electric Vehicle</i>
CALCE	Centro de Engenharia Avançada do Ciclo de Vida	<i>Centre of Advanced Life Cycle Engineering</i>
CNN	Rede Neuronal Convolucional	<i>Convolutional Neural Network</i>
DL	Aprendizagem Profunda	<i>Deep Learning</i>
DNN	Rede Neuronal Profunda	<i>Deep Neural Network</i>
DT	Árvore de Decisão	<i>Decision Tree</i>
EoL	Fim de Vida	<i>End of life</i>
GPR	Regressão por Processo Gaussiano	<i>Gaussian Process Regression</i>
GRU	Unidade Recorrente com Portões	<i>Gated Recurrent Unit</i>
LSTM	Memória de Curto-Longo Prazo	<i>Long Short-Term Memory</i>
MAE	Erro Absoluto Médio	<i>Mean Absolute Error</i>
MAPE	Erro Percentual Absoluto Médio	<i>Mean Absolute Percentage Error</i>
ML	Aprendizagem Computacional	<i>Machine Learning</i>
PdM	Manutenção Preditiva	<i>Predictive Maintenance</i>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

RL	Regressão Linear	<i>Linear Regression</i>
RUL	Vida Útil Restante	<i>Remaining Useful Life</i>
SoH	Estado da Saúde	<i>State of Health</i>
RF	Floresta Aleatória	<i>Random Forest</i>
RMSE	Raiz Quadrada do Erro Médio	<i>Root Mean Squared Error</i>
RNN	Rede Neuronal Recorrente	<i>Recurrent Neural Network</i>
XGBoost	XGBoost	<i>Extreme Gradient Boosting</i>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

INTRODUÇÃO

Contexto

A previsão da vida útil restante (RUL, do inglês *Remaining Useful Life*) de baterias iões de lítio é um ponto fundamental que contribui positivamente para uma estratégia bem planeada de manutenção preditiva (PdM, do inglês *Predictive Maintenance*) nos veículos elétricos, evitando possíveis falhas e avarias. No contexto de baterias, o RUL refere-se ao tempo ou número de ciclos que a bateria ainda pode percorrer até atingir o ponto de fim de vida, que é o ponto em que a capacidade da bateria é igual a 70% relativamente à capacidade inicial.

A degradação da bateria é algo crítico e inevitável, e deve-se a vários fatores, como por exemplo, condições climáticas, ciclos de carga e descarga e a profundidade de carga e descarga. À data, os sistemas de gestão de baterias dos veículos elétricos estão preparados de tal forma que alertam aos utilizadores de possíveis avarias e também de como podem garantir a longevidade da vida das baterias, como por exemplo, o não carregamento das baterias dos veículos até 100% de carga. Nos últimos anos tem-se notado um crescimento substancial da quota de mercado dos veículos elétricos, novos e usados, sendo estes a melhor escolha quando se trata de zero emissões.

Apesar dos aspetos positivos, os veículos elétricos usados já vêm com um histórico de ciclos de carga e de descarga, ou seja, as capacidades de armazenamento de eletricidade das baterias encontram-se abaixo da sua capacidade quando nova, ficando evidente a existência de degradação das baterias.

Motivação e Questões de Investigação

As baterias iões de lítio apresentam diferentes padrões de degradação, mesmo quando são submetidas a condições experimentais controladas, como por exemplo, mesma temperatura ambiente e o mesmo número de ciclos de carga e de descarga. Esta variabilidade tem um impacto direto na robustez da previsão do RUL, em particular na generalização para baterias cujo comportamento de degradação é desconhecido ou até mesmo de baterias novas. Neste âmbito, colocam-se as seguintes questões:

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

1. É possível treinar modelos de aprendizagem computacional com padrões de degradação de baterias conhecidas e adaptá-los para prever o RUL de baterias novas ou de baterias cujo padrões de degradação são desconhecidos?
2. Qual é a capacidade de generalização que os modelos treinados têm em relação a baterias desconhecidas?

Estas questões vão ser analisadas com base no conjunto de dados experimentais *BatteryAgingARC-FY08Q4* da NASA que são amplamente reconhecidos e utilizados na literatura.

Metodologia

A presente dissertação segue a metodologia CRISP-DM (do inglês *Cross Industry Standard Process for Data Mining*) que é composta pelas seguintes etapas: entendimento do projeto; entendimento dos dados; preparação dos dados; modelação; avaliação; e implantação. Estas etapas são fundamentais e têm um grande impacto para o sucesso ou insucesso de um projeto de análise de dados e *data mining*.

A primeira etapa, entendimento do projeto, passa pela definição dos objetivos do estudo no contexto da previsão do RUL de baterias para a PdM de veículos elétricos.

Na etapa do entendimento dos dados, é realizada a análise exploratória dos dados e a descrição do conjunto de dados considerado, concretamente, neste contexto, das características dos perfis de carga e de descarga, e dos padrões de degradação. Nesta etapa procura-se perceber quais são os padrões de degradação das diferentes baterias.

A etapa que segue é referente à preparação dos dados e nela é feito o tratamento e transformações necessárias, nomeadamente: a limpeza dos dados, extração e seleção de variáveis consideradas importantes para a realização da previsão do RUL. Este processo é importante de modo a garantir que os modelos consigam ter um bom desempenho e precisão.

A modelação, etapa quatro, centra-se em treinar, individualmente, os modelos de aprendizagem computacional selecionados para a previsão do RUL. Sendo também feita a otimização dos hiperparâmetros.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Na etapa da avaliação, é realizada a comparação dos modelos a partir das métricas de avaliação de desempenho, comparando os resultados obtidos nas várias iterações. Adicionalmente, é analisada a capacidade de generalização de cada modelo.

Por fim, importa referir que a etapa de implantação será excluída, uma vez que o objetivo deste estudo se limita ao desenvolvimento e análise metodológica e não à implantação operacional.

Estrutura do documento

No que se segue a dissertação está dividida em cinco capítulos e a conclusão. O Capítulo 1 apresenta a revisão da literatura na qual é realizado um enquadramento teórico da manutenção preditiva e do RUL das baterias, com o foco em: descrever as baterias usadas para os veículos elétricos, os modelos orientados por dados e casos de aplicação para a previsão da vida útil restante e o estado da saúde das baterias e por fim aos dados *benchmark*.

De seguida, no Capítulo 2 são apresentados o entendimento do problema e o entendimento dos dados. Este capítulo começa com a definição do objetivo do estudo de uma forma mais detalhada seguindo-se a caracterização do conjunto de dados que será utilizado para a realização da previsão. De modo a complementar o entendimento dos dados, são obtidas as estatísticas descritivas das baterias.

No Capítulo 3, de preparação dos dados, tal como o nome indica, são seguidos os procedimentos de preparação dos dados para a realização da previsão, e estes incluem a limpeza dos dados, extração e seleção de variáveis apenas feita para o conjunto de treino.

Feita a preparação dos dados, o capítulo seguinte é o Capítulo 4, onde são treinados os modelos de aprendizagem computacional que serão utilizados para a previsão do RUL. Ainda neste capítulo, são apresentadas ferramentas de otimização de hiperparâmetros, que serão utilizadas e comparadas a fim de se selecionar a que melhor otimiza os modelos.

No capítulo 5, de resultados, começa-se pela apresentação das métricas de avaliação selecionadas para medir o desempenho dos modelos em cada uma das iterações. De

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

seguida, avaliam-se diferentes técnicas de otimização, selecionando-se a mais eficaz, e, por último, analisam-se comparativamente as três iterações.

Na conclusão, são consolidadas as principais contribuições, destacadas as limitações e são apresentadas propostas para trabalhos futuros.

A opção pela utilização das siglas e acrónimos em inglês está relacionada com o facto de se entender que estas são mais conhecidas na sua versão inglesa, o que facilitará a leitura do texto. Adicionalmente, os dados e *notebooks* usados para o desenvolvimento do trabalho estão disponíveis a partir do link <https://github.com/IbraAlegy/Masters-Thesis>.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

1 REVISÃO DA LITERATURA

Neste capítulo serão abordadas as principais terminologias e apresentados de forma sucinta os resultados obtidos em estudos passados realizados com os conjuntos de dados *benchmark*. A secção inicia com os conceitos de manutenção preditiva, seguido pela vida útil restante das baterias, baterias dos veículos elétricos, os tipos de abordagens utilizadas para a manutenção preditiva destacando a abordagem orientada por dados que é a abordagem aplicada no presente estudo, os modelos tradicionais e modelos mais complexos atualmente utilizados para a manutenção preditiva e por fim os dados *benchmark* usados na literatura.

1.1 Manutenção preditiva

Sem dúvida que uma estratégia de manutenção eficaz é importante para evitar períodos de inatividade e reparações dispendiosas de equipamentos que podem surgir como consequência de um planeamento de manutenção inadequado. A quarta revolução industrial trouxe consigo uma vasta gama de avanços tecnológicos que permitem uma maior eficiência nos processos de produção. Seja a automação, seja a monitorização em tempo real do desgaste com a ajuda da *internet* das coisas e dos sensores que recolhem dados que facilitam a tomada de decisões (Arena et al., 2022).

É importante destacar os diferentes tipos de estratégias de manutenção existentes até à data, começando pela manutenção reativa que é uma estratégia de manutenção em que a intervenção ou reparação apenas ocorre quando o equipamento deixa de funcionar. Por conseguinte, esta estratégia afeta significativamente o tempo de produção, uma vez que o equipamento ficará “fora de serviço” até que o problema seja resolvido, consequentemente, causando graves perdas monetárias (Arena et al., 2022; Hurtado et al., 2023).

De modo a evitar e ultrapassar todos os aspetos negativos de uma estratégia de manutenção reativa, foi introduzida a manutenção preventiva. Esta estratégia consiste em agendar manutenções periódicas do equipamento de modo a evitar possíveis falhas que possam impedir o seu bom funcionamento, minimizando assim a probabilidade de o

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

equipamento ter uma avaria (Achouch et al., 2022). Embora a manutenção preventiva, por um lado, ajude a resolver os problemas da manutenção reativa, por outro lado, podem ser agendadas intervenções sem ter em consideração o estado de saúde atual do equipamento, levando a intervenções de manutenção desnecessárias (Hurtado et al., 2023).

Outra estratégia de manutenção importante que surgiu é a manutenção baseada em condições, que envolve a realização de intervenções baseadas na monitorização das condições atuais do equipamento e automaticamente aciona um alarme antes da ocorrência de uma falha. O principal objetivo da implementação desta estratégia é minimizar os custos de reparação e de inspeção (Nunes et al., 2023).

Como já foi referido, os avanços tecnológicos e a necessidade de ultrapassar todos os aspetos negativos das estratégias de manutenção já existentes, contribuíram para a PdM. Esta foi uma melhoria tão revolucionária que, com a combinação de *internet* das coisas, automação, sensores e análise de dados, a PdM pode prever o ponto em que o equipamento irá falhar, bem como prever a vida útil restante dos equipamentos. A estratégia de PdM envolve uma monitorização contínua permitindo a realização de um planeamento mais preciso e exato da manutenção, melhorando a produtividade e evitando tempos de inatividade (Achouch et al., 2022; Nunes et al., 2023).

De acordo com Nunes et al. (2023), existe uma nova estratégia que leva a manutenção preditiva um passo mais além. Esta estratégia é designada por manutenção prescritiva e utiliza as previsões realizadas pela PdM para dar sugestões sobre formas de como resolver a falha e aumentar o ciclo de vida do equipamento. É importante referir que esta não é uma estratégia simples, uma vez que depende muito das estratégias de manutenção baseada em condições e de PdM bem estabelecidas.

A Figura 1.1 é um resumo ilustrativo da evolução das estratégias de manutenção acima referidas.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

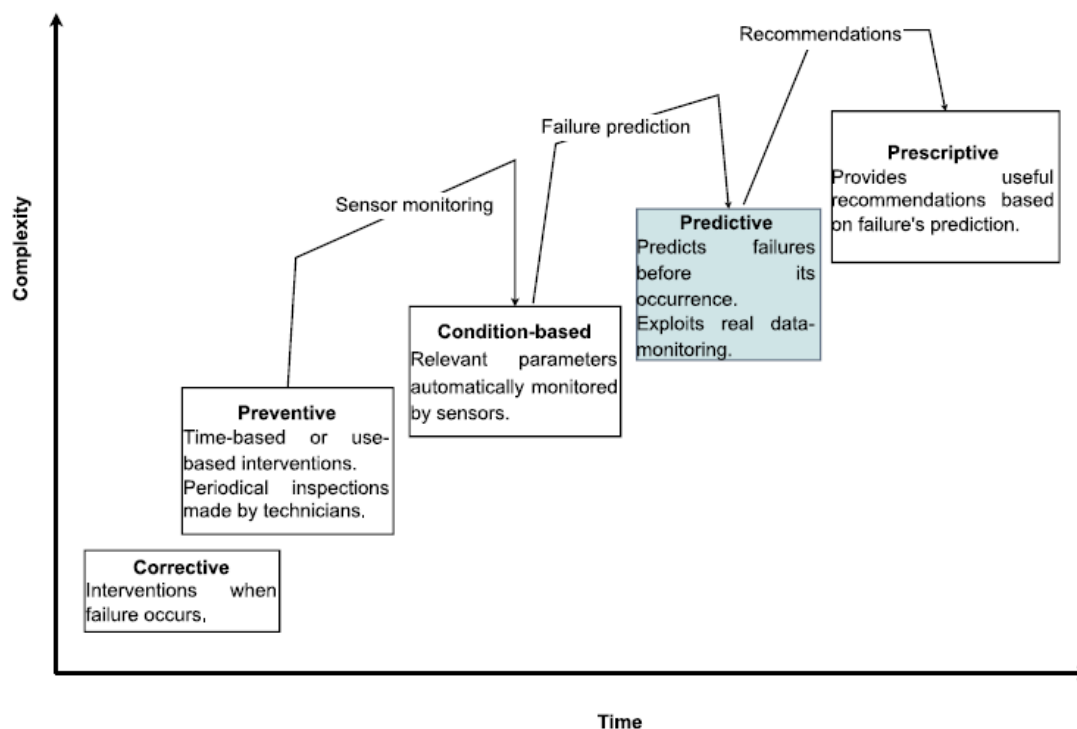


Figura 1.1 Evolução das estratégias de manutenção ao longo do tempo
Fonte: Nunes et al. (2023)

1.1.1 Benefícios da PdM

Importa dar ênfase aos benefícios associados à implementação de PdM destacados por Chang et al. (2023):

1. **Deteção precoce de avarias:** A monitorização contínua permite a deteção precoce de avarias ou anomalias, dando a possibilidade de perceber de que os componentes se estão a degradar a um ritmo fora do normal. O resultado é a prevenção de avarias súbitas e o prolongamento do ciclo de vida do equipamento.
2. **Redução dos custos de manutenção:** A PdM garante a eliminação de agendamentos de manutenção desnecessários e a minimização de falhas através da monitorização constante do estado de saúde do equipamento.
3. **Otimização da utilização dos equipamentos:** Os conhecimentos adquiridos através da PdM permitem que os equipamentos funcionem durante mais tempo a níveis de operação seguros. Tendo em conta que esta estratégia de manutenção

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

pode antecipar a falha do equipamento, as intervenções apropriadas podem ser agendadas nos momentos ideais, minimizando assim o tempo de paragem de produção.

4. **Aumento da segurança e fiabilidade:** A PdM minimiza o risco de falha durante o período crítico de produção ou no funcionamento normal de um equipamento. Por exemplo, nos veículos elétricos, a PdM assegura que os indicadores de desempenho, como o estado de saúde e a capacidade da bateria, se mantêm dentro dos limites de segurança estabelecidos.
5. **Escalabilidade para sistemas complexos:** Tirando partido das técnicas orientadas por dados e com a ajuda da aprendizagem computacional, a PdM tornou-se uma estratégia de manutenção escalável que pode ser aproveitada por várias indústrias. Além disso, a introdução de dados históricos em modelos de aprendizagem computacional pode ajudar a prever com exatidão o estado de saúde e a vida útil restante de um equipamento.

1.1.2 Implementação da estratégia da PdM

A implementação da estratégia de PdM de acordo com Achouch et al. (2022), Arena et al. (2022), Chang et al. (2023), Hurtado et al., (2023) e Nunes et al. (2023), envolve as seguintes etapas essenciais:

1. **Compreender o projeto e recolher os dados relevantes:** Definir objetivos claros da estratégia de PdM e identificar as fontes para recolha de dados.
2. **Processamento e pré-processamento de dados:** Uma etapa crítica para a PdM, que envolve a limpeza, a transformação e a preparação dos dados recolhidos na etapa anterior. Uma boa técnica a utilizar aqui é a deteção de anomalias, que ajuda a identificar e potencialmente remover dados com ruídos, melhorando assim a qualidade dos dados.
3. **Desenvolver e treinar modelos:** Esta etapa envolve a seleção e o desenvolvimento dos modelos de previsão mais adequados, tendo em conta os dados processados.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

4. **Previsão e avaliação:** Os modelos treinados podem ser utilizados para prever o estado de saúde ou a vida útil restante de um equipamento. A avaliação das previsões é fundamental para validar a eficácia do modelo e, posteriormente, tomar decisões de manutenção.
5. **Tomada de decisões e ações de manutenção:** Tal como referido anteriormente, o resultado das previsões ajudará no processo de tomada de decisões de manutenção. O objetivo aqui é programar uma manutenção no período mais adequado.
6. **Implantação e melhoria contínua:** A última etapa consiste em implantar a estratégia de PdM e monitorizá-la continuamente, a fim de a melhorar e integrar novos dados.

1.2 RUL das baterias

É importante reconhecer a necessidade da realização de PdM nas baterias de íões de lítio, em particular nas baterias utilizadas nos veículos elétricos.

Uma noção básica de prognósticos de vida útil é o RUL, que, no fundo, é o tempo ou número de ciclos que faltam até à falha ou avaria de um componente, levando a uma, muito provável, imobilização do equipamento impedindo o seu bom funcionamento (Chen et al., 2021).

Em relação às baterias de íões de lítio Feng et al. (2023), definem RUL como o número de ciclos de carga e descarga a partir do qual se observa que a capacidade utilizável da bateria começa a diminuir até atingir o seu **fim de vida (EoL, do inglês *End-of-Life*)**. O limite do EoL é o ponto em que a capacidade disponível da bateria diminui 20% ou 30% da sua capacidade inicial, ou quando o valor da capacidade cai para 80% ou 70%.

De acordo com Li et al. (2023), o RUL de baterias corresponde ao número de ciclos entre o momento atual e o seu EoL. Este último é definido como o ponto em que a bateria atinge 70% da capacidade nominal inicial, ou seja, quando ocorre uma perda de 30% da capacidade total. Posto isto, o RUL indica quantos ciclos a bateria ainda pode realizar antes de atingir o EoL.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Contudo, e de acordo com Chen et al. (2021), a previsão do RUL é importante para a implementação de estratégias de PdM.

Para além do RUL existe um outro fator importante e considerado na literatura para a PdM. Trata-se do **estado de saúde (SoH)**, do inglês *State of Health*) que caracteriza o estado atual das baterias e vem expresso em percentagem. O SoH de cada bateria é obtido através da avaliação do estado de saúde da bateria de iões de lítio de modo a assegurar um bom funcionamento do sistema de energia e o período possível para a substituição das baterias (Chen et al., 2021).

Zhao et al. (2023) salientam que a previsão exata do RUL e a estimação do SoH das baterias de iões de lítio são fatores importantes do sistema de gestão de baterias. Por sua vez, isto tem um impacto significativo no aumento da vida útil da bateria e na garantia da fiabilidade do sistema de gestão de baterias.

De acordo com Ansari et al. (2021), para garantir uma utilização segura e fiável das baterias, os sistemas de gestão de baterias baseiam-se fortemente na previsão do RUL e na estimativa do SoH. Além disso, as previsões contribuem para o desempenho global do sistema de gestão de baterias, permitindo que o sistema faça uma gestão eficiente dos ciclos de carga e descarga e forneça informações mais precisas sobre o estado atual da bateria e o seu ciclo de vida restante (Chen et al., 2021; Feng et al., 2023).

Em suma, em sistemas de gestão de baterias destacam-se, então, duas áreas de investigação importantes: a previsão do SoH e a previsão do RUL das baterias. A precisão dos métodos de avaliação e previsão tem um impacto direto no desempenho global do sistema de gestão da bateria.

1.3 As baterias nos veículos elétricos

Nos últimos anos, tem-se verificado um aumento notável na aquisição de veículos elétricos a bateria (BEV, do inglês *Battery Electric Vehicles*) devido aos incentivos concedidos pelos governos e à crescente sensibilização para a proteção ambiental.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Tanto nos BEV como nos veículos elétricos híbridos *plug-in*, as baterias de íões de lítio estão na vanguarda no que diz respeito aos sistemas de armazenamento de energia, com uma ênfase crescente na adoção de veículos com emissões baixas ou nulas e na eletrificação dos transportes.

De acordo com Lin et al. (2023), quando as baterias de íões de lítio são comparadas com outras baterias habitualmente utilizadas, como as baterias de níquel-hidrogénio, chumbo-ácido e níquel-crómio, as baterias de íões de lítio são as que mais se destacam no que diz respeito a ciclos de vida mais longos e a uma maior densidade de potência, o que faz deste tipo de baterias um elemento crucial nos sistemas de conversão e gestão de energia dos BEV.

1.3.1 Benefícios de baterias de íões de lítio

Segundo Fanoro et al. (2022), Jia et al. (2023) e Lin et al. (2023), as baterias de íões de lítio oferecem uma vasta gama de vantagens que contribuem para a sua extensiva adoção:

- Como já foi referido, as baterias de íões de lítio têm maior densidade de potência e um ciclo de vida mais longo do que os outros tipos de baterias normalmente utilizados. Isto permite baterias mais leves e compactas para uma necessidade específica de energia e um tempo de vida útil mais longo.
- Uma vez que os BEV exigem uma autonomia substancial e a capacidade de fornecer uma potência elevada para a aceleração, as baterias de íões de lítio demonstram ser capazes de fornecer energia e densidade de potência elevadas.
- Em comparação com outras tecnologias de baterias utilizadas nos BEV, as baterias de íões de lítio têm uma baixa taxa de auto-descarga, o que significa que a bateria mantém a carga durante longos períodos quando o BEV estiver parado ou estacionado.
- As alterações e condições atmosféricas são um dos fatores que são tidos em consideração para os BEV e as respetivas baterias. Por isso, as baterias de íões de lítio provaram funcionar bem numa vasta gama de temperaturas de funcionamento das baterias.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- O equilíbrio é fundamental quando o tema é densidade de potência e energia elevada. Posto isto, as baterias de iões de lítio são capazes de equilibrar a quantidade de energia que pode ser armazenada e a velocidade a que pode ser fornecida.
- As baterias de iões de lítio têm aplicações substanciais e diversificadas nos sistemas energéticos, que vão para além das baterias para BEV e incluem: a capacidade de reserva, a regulação da frequência, o armazenamento e suporte à voltagem. O facto de as baterias de iões de lítio serem versáteis realça a sua importância nas infraestruturas energéticas modernas.

1.3.2 Características das baterias de iões de lítio

De acordo com Fanoro et al. (2022), uma bateria de iões de lítio é composta por dois elétrodos (ânodo e cátodo) e um eletrólito, cada um contendo uma pequena quantidade de lítio móvel. Durante os ciclos de carga e descarga, os iões de lítio são extraídos de um elétrodo, migram através do eletrólito e são depois introduzidos no outro elétrodo, enquanto os eletrões circulam entre os elétrodos através de um circuito externo. Este processo, anteriormente descrito, é considerado essencial para o funcionamento da bateria. A Figura 1.2 ilustra visualmente o processo de descarregamento da bateria.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

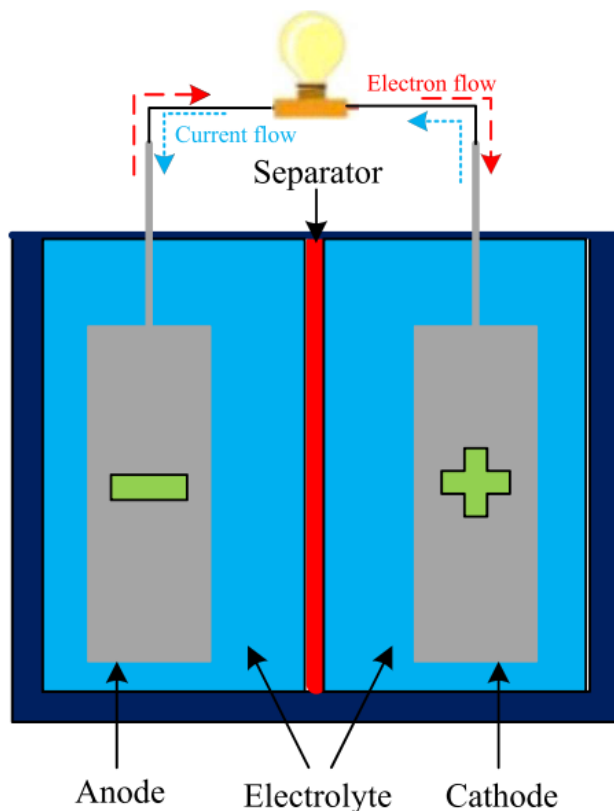


Figura 1.2 Célula de iões de lítio em descarregamento. Importa referir que o processo de carregamento ocorre no oposto.

Fonte: Ahmadian et al. (2018)

Segundo Ezzouhri et al. (2023) e Fanoro et al. (2022), o desempenho das baterias nos BEV pode determinar significativamente as limitações e a fiabilidade do veículo. Posto isto, as baterias de iões de lítio têm várias características-chave que realçam tanto as limitações como a fiabilidade:

- **Deterioração do desempenho ao longo do tempo** devido a alterações físicas e químicas irreversíveis que ocorrem durante a utilização regular das baterias, por exemplo, ciclos de carga e descarga, ou mesmo armazenamento.
- Os fatores que desempenham um papel significativo no estado de degradação da bateria incluem a **frequência**, os **ciclos de descarga**, a **procura de corrente**, a **profundidade da carga e da descarga** e as **variações na temperatura de funcionamento**.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- A **perda de capacidade** ocorre quando a capacidade da bateria diminui gradualmente ao longo do tempo até a bateria atingir o seu limite de EoL.
- Também é importante destacar aqui os dois tipos de degradação das baterias, a **degradação do calendário** e a **degradação do ciclo**. A degradação do calendário refere-se à perda de capacidade durante o armazenamento e a degradação do ciclo, como o nome sugere, ocorre devido aos ciclos de carga e descarga.

1.3.3 Degradação do ciclo das baterias

Tal como mencionado no último ponto da subsecção anterior, a degradação do ciclo refere-se à degradação e à diminuição do desempenho que ocorre com cada ciclo de carga e descarga. Este tipo de degradação é caracterizado principalmente pela perda de capacidade e por um aumento da resistência interna ao longo do tempo de vida da bateria (Lin et al., 2023).

De acordo com Atalay et al. (2020), Ezzouhri et al. (2023), Fanoro et al. (2022) e Lin et al. (2023), os principais aspetos da degradação do ciclo são os seguintes:

- À medida que o número de ciclos de carga e descarga aumenta, o desempenho das baterias de íões de lítio diminui gradualmente.
- A degradação da bateria é inevitável e é causada por uma variedade de fatores, incluindo o armazenamento, o comportamento na condução e os ciclos de carga e descarga. O ritmo a que a bateria se degrada varia de acordo com os fatores anteriormente mencionados. Por exemplo, se o veículo ficar parado durante um período com 100% de carga armazenada, isto pode contribuir para acelerar o processo de degradação da bateria.
- A degradação a nível do material inclui a decomposição estrutural do cátodo (elétrodo positivo), a formação de camadas de interfase eletrolítica no ânodo (elétrodo negativo), a fratura das partículas do ânodo em resultado das tensões de inserção e extração do lítio e a amontoação de lítio metálico na superfície do ânodo em resultado de determinadas condições de carga.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- Durante os ciclos (carga/descarga), a tensão induzida pela difusão associada à intercalação de iões de lítio provoca a deformação do material do eletrodo, a pulverização, a fissuração e a fratura. Como resultado, há um aumento da resistência elétrica e do isolamento do material ativo, contribuindo diretamente para a diminuição da capacidade e, eventualmente, para a falha da bateria.
- A utilização intensiva e frequente da bateria conduz a uma deterioração acelerada do desempenho da bateria. Por outras palavras, taxas de carga elevadas e temperaturas de carga baixas podem agravar a degradação, o que pode potencialmente levar a problemas como o revestimento de lítio.
- Um parâmetro crucial que está altamente correlacionado com a capacidade da bateria e serve como indicador da tendência geral de degradação da bateria é a resistência interna. Posto isto, um aumento neste parâmetro é uma indicação de degradação da bateria.

De um modo geral, a compreensão dos aspetos de degradação do ciclo acima mencionados é fundamental para o desenvolvimento de métodos fiáveis e precisos de estimativa de SoH e, assim, para a construção eficaz de um sistema de gestão de baterias que possa garantir um funcionamento seguro, otimizar a vida útil da bateria e fornecer uma previsão fiável do RUL das baterias de iões de lítio (Atalay et al., 2020; Lin et al., 2023).

1.4 Abordagem orientada por dados

A implementação da PdM, como já referido, passa pelo processo de recolha e interpretação de dados adquiridos através de sensores, e isto torna-se possível com recurso a técnicas de inteligência artificial e aprendizagem computacional (ML, do inglês *Machine Learning*) (Arena et al., 2022).

A abordagem orientada por dados para PdM utiliza dados históricos e em tempo real extraídos de sensores e de outras fontes para estimar o SoH e prever o RUL de um equipamento ou dos seus componentes (Arena et al., 2022; Hurtado et al., 2023; Zhao et al., 2023). Ao contrário das abordagens orientadas por modelos físicos, que são mais

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

complexas e precisam do conhecimento técnico que permite descrever o comportamento e degradação de um equipamento, as abordagens orientadas por dados baseiam-se em dados do equipamento para aprender padrões e relações sem exigir um conhecimento profundo dos mecanismos de falha subjacentes (Lin et al., 2023; Sharma & Bora, 2023; Zhu et al., 2023).

1.4.1 Abordagens baseadas em ML

Segundo Çinar et al. (2020), ML é uma subárea da inteligência artificial e é entendido como qualquer algoritmo ou programa que demonstra capacidade de aprender, com menor ou nenhum suporte adicional, a partir de dados históricos.

As abordagens de ML são, usualmente, divididas em quatro tipos de aprendizagem, nomeadamente, não supervisionada, semi-supervisionada, supervisionada e por reforço, que Theissler et al. (2021) explicam da seguinte forma:

- Na **aprendizagem não supervisionada**, os dados não são etiquetados, e o modelo é formulado de forma a identificar, por si, padrões e estruturas nos dados.
- Relativamente à **aprendizagem semi-supervisionada**, a característica dos dados introduzidos é uma combinação de dados etiquetados e não etiquetados.
- Para a **aprendizagem supervisionada**, o modelo de ML utiliza dados de formação etiquetados. Por outras palavras, são atribuídas etiquetas e o modelo tem como objetivo aprender um mapeamento das entradas para as saídas, ajustando frequentemente o modelo de uma forma iterativa, sendo este um processo que é repetido de modo a atingir o nível desejado de precisão e de modo a poder prever corretamente os resultados.
- A **aprendizagem por reforço** permite que o sistema aprenda por regra de tentativa e erro de modo a descobrir as ações mais benéficas. As aplicações da aprendizagem por reforço na indústria automóvel têm sido fundamentais permitindo o desenvolvimento de veículos com capacidade de autocondução que aprendem a reconhecer o ambiente envolvente e que adaptam o seu comportamento com base nas diversas situações que possa enfrentar.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Na indústria automóvel, os algoritmos de ML exigem uma análise eficaz de grande quantidade de dados históricos e de dados em tempo real extrapolados de sensores e sistemas de tecnologia de informação. Posto isto, Arena et al. (2022) alertam que a fase de pré-processamento dos dados tem um grande impacto no desempenho dos algoritmos de ML.

Na literatura, vários modelos de ML são normalmente utilizados para realizar previsões no contexto de PdM, prognóstico, diagnóstico de avarias, estimar SoH e RUL. Apresentam-se de seguida algumas técnicas tradicionais de ML supervisionadas utilizadas no contexto da previsão de PdM:

- A **regressão linear (RL)** é um modelo tradicional de ML que é utilizado para fazer a representação da conexão entre uma variável dependente e uma ou várias variáveis independentes. O pressuposto deste modelo é a existência de uma relação linear entre variáveis, em que no caso de alterações na variável dependente variam de forma diretamente proporcional nas variáveis independentes. A RL é utilizada para a PdM como um modelo base para comparar o resultado com outros modelos de ML (Arena et al., 2022; Jafari et al., 2024).
- Os ***k*-vizinhos mais próximos** (do inglês *k-Nearest Neighbors*) é um algoritmo utilizado principalmente para a classificação de falhas e o reconhecimento de padrões, que classifica as novas instâncias com base na classe maioritária entre os seus *k* vizinhos mais próximos dos dados de treino. Embora este algoritmo não seja o mais adequado para a previsão de RUL, tem sido utilizado em conjunto com outros (Hurtado et al., 2023).
- **Máquina de vetor de suporte** (em inglês *Support Vector Machine*) é um modelo que funciona bem na procura de um limite de decisão ótimo, que separa os pontos de dados que pertencem a classes diferentes. Por outro lado, a regressão por vetor de suporte (em inglês *Support Vector Regression*) é utilizado para problemas de regressão, cujo objetivo é encontrar uma função que melhor se ajusta aos dados, minimizando o erro entre as previsões e os valores reais.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Estudos anteriores aplicaram estes modelos para diagnosticar falhas, classificar tarefas e prever RUL em PdM (Hurtado et al., 2023; Sharma & Bora, 2023; Theissler et al., 2021).

- A **regressão por processo gaussiano (GPR, do inglês *Gaussian Process Regression*)** é um modelo bayesiano não paramétrico utilizado para fins de classificação e regressão (Sharma & Bora, 2023). Esta técnica de ML é uma ferramenta eficiente para incorporar conhecimento prévio e desenvolver modelos de previsão. Além disso, em estudos anteriores, o GPR foi utilizado para a previsão de RUL e demonstrou ter um bom desempenho em comparação com outros modelos de ML (Arena et al., 2022).
- As **árvores de decisão (DT, do inglês *Decision Tree*)** são também métodos não paramétricos que aprendem regras de decisão a partir de dados. As DT têm sido utilizadas em cenários de diagnóstico e prognóstico de falhas com resultados positivos. A velocidade relacionada ao treino de dados e de previsão utilizando DT tem muitas vantagens para problemas em tempo real (Hurtado et al., 2023).
- As **florestas aleatórias (RF, do inglês *Random Forest*)** são técnicas de ML em conjunto que combinam em paralelo várias DT para melhorar a precisão de uma previsão, por outras palavras, o algoritmo RF calcula a média e integra os resultados de muitas DT. Com a agregação de previsões de DT individuais, o modelo proporciona uma maior precisão na identificação de relações complexas que existem dentro dos dados. A utilização deste algoritmo tem apresentado resultados positivos na realização da deteção de avarias (Arena et al., 2022; Jafari et al., 2023).
- O **XGBoost (em inglês *Extreme Gradient Boosting*)** é um algoritmo de ML que tem como premissa o treino sequencial das DT de modo a captar padrões complexos que existem nos dados. Cada árvore criada pelo algoritmo, mapeia os dados de entrada, de forma aleatória, para os diferentes nós gerando assim uma pontuação contínua para cada um dos nós. A estimação dos dados de treino, resíduos ou erros correspondem à diferença entre o real e os pontos de dados previstos. O treino do XGBoost ocorre de forma iterativa, em que as árvores mais

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

recentes são criadas tendo em consideração o erro ou resíduo da árvore anterior. Este algoritmo é uma escolha popular para a previsão do RUL devido a sua adaptabilidade (Mohril et al., 2023; Sharma & Bora, 2023).

- **CatBoost** é uma técnica de ML moderna, mais apta a lidar com variáveis categóricas, e baseia-se em reforço gradiente de DT que demonstrou uma elevada precisão em tarefas de previsão (Sharma & Bora, 2023).
- O **AdaBoost.Rt** funciona como um algoritmo de reforço que cria um forte modelo de previsão através da construção sequencial de uma série de modelos que, inicialmente, são “modelos fracos”. O modelo ajusta os pesos ou a importância dos pontos de dados de treino. Além disso, os pontos de dados que foram mal previstos recebem posteriormente pesos mais elevados. Contudo, o modelo AdaBoost.Rt reúne todos os “modelos fracos” individuais para criar com mais precisão um único “modelo forte” (Li et al., 2022).

1.4.2 Abordagens baseadas em aprendizagem profunda

A aprendizagem profunda (DL, do inglês *Deep Learning*) é uma subárea de ML, incorporando algoritmos capazes de aprendizagens supervisionadas e não supervisionadas (Achouch et al., 2022; Arena et al., 2022). DL utiliza redes neurais artificiais com várias camadas de modo a estimar uma melhor função de mapeamento entre determinados dados e resultados (Arena et al., 2022; Chen et al., 2021). As técnicas de DL podem ser aplicadas a equipamentos industriais para a realizar previsões de falhas, deteção de falhas e entre outros (Achouch et al., 2022).

Apresentam-se de seguida algumas técnicas tradicionais de DL utilizadas no contexto de PdM:

- As **redes neurais artificiais** (ANN, do inglês *Artificial Neural Network*) são amplamente utilizadas na PdM para a identificação, deteção e previsão de falhas (Nunes et al., 2023). As ANN são capazes de aprender relações complexas e não lineares a partir de dados e, por sua vez, proporcionam uma elevada exatidão e velocidade de computação, tudo isto sem exigir um conhecimento prévio das

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

caraterísticas físicas do sistema (Elmahallawy et al., 2022). Relativamente às baterias, as ANN têm demonstrado um bom desempenho na estimação do SoH (Arena et al., 2022).

- As **redes neuronais recorrentes (RNN**, do inglês *Recurrent Neural Network*) são especificamente concebidas para trabalhar com dados sequenciais, tornando o modelo adequado para prever falhas ao longo do tempo (Hurtado et al., 2023). Um tipo de RNN é a **memória de curto-longo prazo (LSTM**, do inglês *Long Short-Term Memory*), que é uma rede concebida para lidar com dependências de longo alcance em dados de séries temporais e tem sido utilizada para prognósticos, como RUL e SoH (Achouch et al., 2022; Arena et al., 2022). Além disso, para melhorar a precisão da previsão, alguns autores utilizaram uma abordagem híbrida, combinando o modelo LSTM com outras técnicas.
- As **redes neuronais convolucionais (CNN**, do inglês *Convolutional Neural Network*) funcionam bem na pesquisa de padrões locais em dados estruturados e podem também ser aplicadas a dados de séries temporais unidimensionais para prever a degradação ao longo do tempo e para a deteção de falhas (Achouch et al., 2022; Hurtado et al., 2023). As CNN têm sido utilizadas para a deteção de falhas em vários sensores (Arena et al., 2022).
- As **redes generativas adversariais** são uma técnica de DL utilizada principalmente para gerar dados sintéticos com o objetivo de tratar conjuntos de dados desequilibrados em PdM e podem também ser utilizadas para estimar indicadores de saúde e para a deteção de anomalias (Hurtado et al., 2023).
- Os **autoencoders** são redes neuronais treinadas para aprender uma representação comprimida dos dados de entrada. Além disso, os *autoencoders* são utilizados para a deteção de anomalias e a extração de caraterísticas em problemas de prognóstico e diagnóstico de falhas (Hurtado et al., 2023).

1.4.3 Casos de aplicação para a previsão do SoH

Li et al. (2023) propõem uma abordagem orientada por dados para estimar o SoH que integra caraterísticas de saúde externas e internas da bateria. Os autores utilizam ANN e

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

RF para estimar o SoH. Além disso, os autores destacam que as características de saúde internas e externas demonstram um desempenho superior ao estimar o SoH *offline* e *online*, apresentando assim uma estimativa mais robusta, mesmo em condições de ruído.

Lin et al. (2023) apresentam uma abordagem diferente para estimar o SoH que tem em conta a resistência interna da bateria. O método proposto utiliza o *Explanation Boosting Machine*, que facilita a interpretação e o desempenho do ajuste. Ao comparar esta abordagem com outras, como o LSTM e o GPR, o modelo parece ser mais robusto e tem uma precisão melhorada na estimativa do SoH.

Jafari et al. (2024) propõem um novo método de previsão da perda de capacidade da bateria utilizando características como a voltagem, a temperatura e a corrente. Os autores utilizam vários tipos de modelos de ML, partindo de um modelo menos complexo, precisamente a RL, até modelos mais complexos, como o XGBoost e o RF. Posto isto, o modelo proposto foi baseado num XGBoost otimizado, centrado na otimização dos parâmetros do XGBoost. Comparando os resultados de todos os modelos utilizados no estudo, o modelo XGBoost otimizado proposto pelos autores superou todos os outros modelos na previsão da perda da capacidade das baterias de íões de lítio.

Wu et al. (2023) propõem um método de estimação de SoH que se destaca pela sua capacidade de desempenho eficiente em diferentes tipos de conjuntos de dados de baterias incluindo conjuntos de dados dos repositórios CALCE, NASA e Oxford. O método proposto utiliza três modelos, nomeadamente LSTM, GPR e rede neuronal de retro propagação. Os resultados obtidos mostram que este último, o modelo rede neuronal de retro propagação, em comparação com os outros dois modelos, tem a maior precisão de estimativa quando se trata de estimar o SoH de uma bateria.

Mantendo o foco em baterias, Li et al. (2022) propõem o modelo AdaBoost.Rt-RNN para estimar o estado de carga, combinando reforço e RNN com um conjunto diversificado de variáveis. Os resultados obtidos indicam claramente que o modelo proposto consegue alcançar uma maior precisão ao estimar o estado de carga quando comparado com uma única RNN e outros modelos *benchmark*, como o XGBoost e a rede neuronal de retro propagação utilizados no estudo.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

A Tabela 1.1 apresenta de forma resumida os casos de aplicação de modelos de ML e/ou de DL para a previsão do SoH. A coluna dados serve para indicar a fonte dos conjuntos de dados usados nos casos de aplicação acima referidos. Os conjuntos de dados serão apresentados com mais detalhes na Secção 1.5.

Tabela 1.1 Casos de aplicação de modelos de ML e/ou de DL para a previsão do SoH.

Autores	Ano	Objetivo	Modelos	Dados
Li et al.	2022	Um modelo híbrido para estimar o estado de carga.	AdaBoost.Rt, RNN	-
Lin et al.	2023	Uma abordagem para estimar o SoH tendo em conta a resistência interna da bateria	LSTM, GPR, SVR	Oxford
Li et al.	2023	Uma abordagem orientada por dados para estimar o SoH que integra características de saúde externas e internas da bateria	ANN, RF	-
Wu et al.	2023	Uma abordagem orientada por dados para estimar o SoH	LSTM, GPR, BPNN	CALCE, NASA, Oxford
Jafari et al.	2024	previsão da perda de capacidade com recurso a XGBoost otimizado	RL, RF, XGBoost	Elsevier Scopus

Legenda: BPNN – rede neuronal de retro propagação, SVR – regressão por vetores de suporte.

1.4.4 Casos de aplicação para a previsão do RUL

Feng et al. (2023) propõem um modelo híbrido orientado por dados, combinando modelos para prever o RUL de baterias de iões de lítio. Os autores utilizam a análise de componentes principais para reduzir a complexidade do modelo, encurtando a dimensão das características extraídas. Depois deste processamento, são introduzidas no modelo híbrido composto por CNN e LSTM bidirecional, capturando assim as dependências temporais nos dados de degradação.

Wang et al. (2023) introduzem uma nova forma de realizar a previsão do RUL de baterias de iões de lítio baseada na abordagem *complete ensemble empirical mode decomposition with adaptive noise* e um *transformer*. Com a utilização do modelo, os autores conseguem decompor a sequência de capacidade bruta em flutuações de curto prazo e um resíduo de longo prazo. Posteriormente, com a utilização de um modelo *transformer*, os autores

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

tratam as dependências de longo alcance nos dados de perda de capacidade, com o objetivo de melhorar a precisão da previsão.

Liu et al. (2023) propõem um modelo de características de intervalos de pico e LSTM para prever o RUL de baterias de íões de lítio. Ao utilizar este modelo, os autores extraem características baseadas em intervalos de pico dos dados da bateria, posteriormente, as características extraídas são treinadas utilizando o modelo LSTM. Além disso, o modelo proposto tem impactos de previsão superiores quando comparado com diferentes modelos DL e outros modelos ML utilizando os mesmos conjuntos de dados.

Jafari et al. (2023) propõem uma abordagem com recurso a um modelo híbrido combinando os algoritmos KNN, RF e XGBoost para prever o RUL e a perda de capacidade de baterias íões de lítio. Para além de demonstrar os resultados obtidos com o modelo híbrido, os autores também realizaram a previsão com cada um dos modelos e no final fizeram a avaliação comparativa do desempenho dos quatro modelos de previsão.

Ezzouhri et al. (2023) descrevem uma abordagem orientada por dados para a previsão do RUL de baterias de íões de lítio, que utiliza modelos ML incluindo RL, RF e *CatBoost*. Os autores usam uma estrutura baseada em uma nova estratégia de extração de variáveis que tem como objetivo melhorar as previsões. Esta nova extração de variáveis está dividida em três grupos: variáveis do estado inicial, variáveis do estado atual e variáveis do estado histórico. Os resultados experimentais indicam que todos os modelos têm um bom desempenho com os conjuntos de dados *benchmark* de baterias da NASA.

Ansari et al. (2021) propõem uma previsão do RUL de bateria de íões de lítio orientada por dados com recurso ao modelo RNN com um perfil de entrada multicanal. Primeiro, o perfil de entrada multicanal gera um formato de dados de entrada de 31 canais com a incorporação de vários parâmetros dos perfis de carregamento da bateria, incluindo temperatura, voltagem, corrente, capacidade e descarga. Posteriormente, o modelo RNN é utilizado para processar os dados de entrada multicanal e prever o RUL.

Chang et al. (2023) utilizam quatro modelos para prever o RUL de baterias de íões de lítio e comparam o desempenho de cada modelo tendo em conta as variáveis de carga e

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

descarga extraídas do conjunto de dados de baterias da NASA. Os modelos incluem DT, RNN, *nonlinear autoregressive with external input* e LSTM. Dos quatro modelos utilizados por estes autores, o LSTM é o modelo com melhor desempenho em termos de erro de previsão.

Chen et al. (2021) propõem uma abordagem orientada por dados dividida em duas fases, uma para a previsão de SoH e outra para a previsão de RUL de baterias de íões de lítio, com recurso a modelos de DL. Os modelos baseiam-se na *empirical mode decomposition*, RNN, LSTM e *Gated Recurrent Unit* (GRU). Em ambas as fases, foram efetuadas previsões para cada bateria, avaliando o desempenho de cada modelo em diferentes pontos de partida.

Chen et al. (2022) propõem uma abordagem híbrida de prognóstico integrado que inclui uma Decomposição de Sequências e DL para prever o RUL de baterias de íões de lítio. Esta abordagem pré-processa os dados da bateria combinando a *complementary ensemble empirical mode decomposition* e a análise de componentes principais para reduzir a dimensionalidade e separar os efeitos da regeneração da capacidade. A abordagem proposta separa os diferentes comportamentos de degradação, a fim de melhorar a precisão da previsão.

Li et al. (2023) propõem um modelo híbrido orientado por dados, combinando modelos para prever o RUL de baterias de íões de lítio. O modelo proposto utiliza uma estrutura *encoder-decoder*, em que a rede convolucional temporal é utilizada para o módulo codificador e o GRU é utilizado para o módulo decodificador.

A Tabela 1.2 apresenta de forma resumida os casos de aplicação de modelos de ML e/ou de DL para a previsão do RUL.

Tabela 1.2 Casos de aplicação de modelos de ML e/ou de DL para a previsão do RUL.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Autores	Ano	Objetivo	Modelos	Dados
Ansari et al.	2021	Uma abordagem com perfil de entrada multicanal para a previsão do RUL de baterias	RNN	NASA
Chen et al.	2021	Uma abordagem híbrida dividida em duas fases: 1ª Previsão do SoH, 2ª RUL de baterias	RNN, LSTM, GRU	NASA
Chen et al.	2022	Uma abordagem híbrida que com recurso a Decomposição de Sequências e DL Previsão do RUL de baterias	CEEMD, LSTM	CALCE, NASA
Chang et al.	2023	Uma abordagem orientada por dados para a previsão do RUL de baterias	DT, RNN, NARX, LSTM	NASA
Ezzouhri et al.	2023	Uma abordagem orientada por dados para a previsão do RUL de baterias	RL, RF, CatBoost	NASA
Feng et al.	2023	Um modelo híbrido para previsão do RUL de baterias utilizando PCA	CNN, LSTM	CALCE, NASA
Jafari et al.	2023	Um modelo híbrido para previsão do RUL e a perda de capacidade das baterias	KNN, RF, XGBoost	NASA
Li et al.	2023	Um modelo híbrido que utiliza uma estrutura <i>Encoder-Decoder</i> para previsão do RUL	TCN, GRU, DNN	CALCE, NASA
Liu et al.	2023	Um modelo de características de intervalos de pico e LSTM para prever o RUL de baterias	LSTM	NASA
Wang et al.	2023	Previsão do RUL das baterias baseada na CEEMDAN	CEEMDAN	CALCE, NASA

Legenda: CEEMD - *Complementary Ensemble Empirical Mode Decomposition*, CEEMDAN - *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise*, GRU – unidade recorrente com portões, NARX - *nonlinear autoregressive with external input*, TCN - rede convolucional temporal, DNN – rede neuronal profunda.

1.5 Dados *benchmark*

A *internet* contribuiu para o aumento drástico da disponibilidade de dados que podem também ser utilizados para gerar dados simulados de *software*. Ao longo dos anos, vários conjuntos de dados têm sido publicados devido o crescente interesse pela investigação em tecnologia de baterias, em particular as baterias de íões de lítio. Os conjuntos de dados encontram-se disponíveis *online*, sendo que a maioria pode ser utilizada para a realização

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

de investigação sob algumas condições. Os conjuntos de dados de baterias são construídos com recurso a uma variedade de modelos de baterias e podem ser úteis para realizar a previsão de muitos estados de baterias como, por exemplo, o SoH, estado de carga, estado de segurança, estado de energia e estado de temperatura, que podem ser úteis no apoio à melhor gestão da bateria e à realização da previsão do RUL e a estimação do SoH (Hasib et al., 2021).

Nos casos de aplicação de modelos para a previsão do RUL e para a previsão do SoH que foram apresentados na secção anterior, destacam-se dados provenientes do repositório da National Aeronautics and Space Administration (NASA), em particular o conjunto de dados *BatteryAgingARC-FY08Q4*¹, composto por quatro baterias do tipo iões de lítio, nomeadamente, B0005, B0006, B0007 e B0018. O conjunto de dados foi recolhido a partir de um banco de testes de prognósticos de baterias no Centro de Excelência em Prognóstico da NASA Ames. As baterias foram submetidas a três perfis operacionais, nomeadamente, carga, descarga e impedância e a temperatura ambiente de 24°C, permitindo registar a degradação das mesmas ao longo da experiência (Saha & Goebel, 2007).

Para além do repositório da NASA, existem alternativas na procura de dados de baterias publicamente disponíveis *online*, como por exemplo os repositórios da CALCE e IEEE.

O Centre of Advanced Life Cycle Engineering (CALCE) tem no seu repositório um conjunto de dados de baterias do tipo iões de lítio, como por exemplo CS2, que também são testados em várias condições de temperatura ambiente e ciclos de carga e descarga. Este conjunto de dados inclui as baterias CS2-35, CS2-36, CS2-37 e CS2-38 (Hasib et al., 2021).

O Portal de dados do Institute of Electrical and Electronics Engineers (IEEE) fornece um conjunto de dados simulados a partir da utilização de uma célula de bateria de polímero de lítio modelo *ePLB C020* num veículo elétrico. Os dados já se encontram divididos em

¹ Disponível em <https://phm-datasets.s3.amazonaws.com/NASA/5.+Battery+Data+Sct.zip>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

conjuntos de treino e de teste, e podem ser utilizados para estimar o estado de carga e o SoH da bateria (Reis et al., 2021).

É importante mencionar aqui que, na literatura, o conjunto de dados mais utilizado para prever o RUL e para estimar a perda de capacidade ou o SoH são os conjuntos de dados da NASA e do CALCE. Alguns artigos utilizam estes dois conjuntos de dados disponíveis publicamente para analisar e comparar o desempenho dos modelos propostos quando utilizados em conjuntos de dados diferentes.

De seguida são apresentados alguns casos de uso dos conjuntos de dados de NASA e CALCE.

Chen et al. (2021) utilizam conjuntos de dados da *BatteryAgingARC-FY08Q4*, para fazer a previsão do SoH e do RUL de cada uma das baterias. Por outras palavras, utilizaram os dados históricos de cada uma das baterias para fazer a previsão dividindo os conjuntos de treino, validação e teste em 40/30/30, 60/20/20 e 80/10/10.

Ansari et al. (2021) apresentam duas abordagens, nomeadamente, uma abordagem unicanal e uma abordagem multicanal para a previsão do RUL das baterias B0005, B0006, B0007 e B0018. Para a abordagem unicanal, os autores utilizam os dados históricos de cada bateria para prever o RUL de cada conjunto de dados. Por outro lado, a abordagem multicanal, permitiu-lhes realizar a previsão do RUL de uma bateria, experimentando várias combinações, ou seja, consideram para o conjunto de treino a combinação dos dados de duas baterias e de três baterias e para o conjunto de teste consideram os dados de uma bateria.

Numa abordagem similar, Jafari e Byun (2022) utilizam os dados históricos da degradação das baterias B0005, B0006, B0007 para prever o RUL da bateria B0018.

2 ENTENDIMENTO DO PROBLEMA E DOS DADOS

Neste capítulo serão apresentados o entendimento do problema e dos dados que são fundamentais para o estudo, dando assim mais contexto ao problema e à apresentação dos conjuntos de dados que serão utilizados para o estudo.

2.1 Entendimento do problema

A crescente utilização de baterias de íões de lítio para BEV tornou a previsão do RUL uma tarefa fundamental para garantir a segurança e a fiabilidade das baterias. Conforme mencionado anteriormente, RUL refere-se ao número de ciclos necessários para que uma bateria atinja o seu EoL, ou seja, quando a bateria atinge o limite de perda de capacidade, concretamente quando a capacidade da bateria desce para os 70%.

A previsão do RUL das baterias contribui para a PdM dos BEV, maximizando assim a utilização da bateria e contribuindo para evitar falhas repentinas. Embora isso seja verdade, é importante compreender que podem existir desafios quando se trata da implementação de modelos em diferentes baterias, porque cada uma tem características distintas, como por exemplo, padrões de degradação e capacidade inicial que pode ser diferente da capacidade nominal.

Nesta dissertação, o problema específico que é abordado e tratado é a previsão do RUL de uma bateria com base nos dados históricos de outras baterias.

A contribuição desta dissertação consiste em analisar como os modelos clássicos de ML, nomeadamente RL, RF e XGBoost, podem ser treinados com dados históricos de baterias conhecidas e adaptados para prever o RUL de novas baterias.

Por fim, refira-se que o conjunto de dados usado é *BatteryAgingARC-FY08Q4* da NASA, e foram selecionados por serem dados frequentemente utilizados na literatura, tal como se pode constatar nos casos de aplicação apresentados na Capítulo 1.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

2.2 Entendimento dos dados

O conjunto de dados usado é composto por quatro baterias recarregáveis do tipo íões de lítio 18650 que foram submetidos a ciclos de carga, descarga e impedância à temperatura ambiente de 24°C. O perfil de carregamento foi efetuado com a corrente constante de carga a 1,5A até que a voltagem atinja 4,2V seguido de uma voltagem constante até que a corrente de carga desça para 20mA. O perfil de descarga para as baterias B0005, B0006, B0007 e B0018 foi efetuado com a corrente constante de descarga a 2A até que a voltagem desça para 2,7V, 2,5V, 2,2V e 2,5V, respetivamente. O perfil de impedância permite obter informação do diagnóstico sobre os processos internos da bateria. Apesar deste perfil fazer parte deste conjunto de dados, não é informação que esteja sempre presente neste contexto de previsão associado a este tipo de baterias, pelo que foram considerados apenas os perfis de carga e de descarga.

Segundo a descrição do conjunto de dados, as experiências foram interrompidas quando as baterias atingiram o EoL concretamente, quando se registou uma diminuição de 30%, ou seja, de 2Ah para 1,4Ah, da capacidade nominal de cada bateria. Embora o conjunto de dados tenha essa informação, é possível observar na Figura 2.1 que, à exceção da bateria B0007, todas as baterias atingem o EoL e a experiência continua por mais alguns ciclos, o que dá a entender que os testes não foram suspensos imediatamente após as baterias perderem 30% da capacidade nominal.

Contudo, e uma vez que se pretende utilizar as quatro baterias, sabendo que a bateria B0007 não chega a atingir o EoL foi considerado 29% de diminuição da capacidade nominal o que representa a 1,41Ah.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

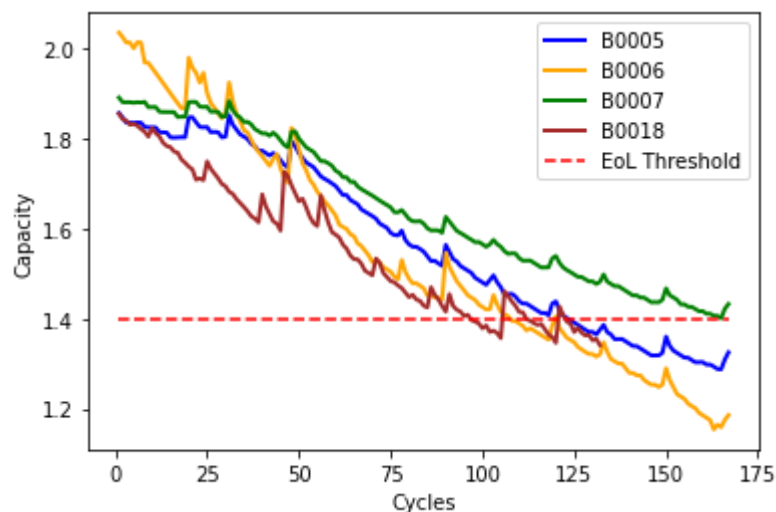


Figura 2.1 Degradação da capacidade das baterias

As baterias B0005, B0006 e B0007 têm 170 ciclos de carga e 168 ciclos de descarga. A bateria B0018, conta com 134 ciclos de carga e 132 ciclos de descarga. Cada ciclo de carga e de descarga é composto por vários registos que relatam o ponto inicial do ciclo até o ponto em que o ciclo termina. A Tabela 2.1 contém informação presente em cada um dos ciclos e a respetiva descrição. Importa referir que cada ciclo de carga e de descarga contém um número de registos que refletem informação recolhida desde o início do ciclo até o fim do mesmo, sendo que este número varia de ciclo para ciclo.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Tabela 2.1 Informação existente em cada um do tipo ciclo.

Ciclos	Atributos	Descrição
Carga e descarga	Ciclo	Ciclos de carga e de descarga
	Tipo	O perfil do ciclo, ou seja, o perfil de carga e o perfil de descarga
	Temperatura ambiente	A temperatura usada durante a experiência
	<i>Timestamp</i>	A data e hora do início de cada ciclo de carga e de descarga
Carga	Voltagem medida	Voltagem do terminal da bateria medida em Volts
	Corrente medida	Corrente de saída da bateria medida em amperes
	Temperatura medida	Temperatura da bateria medida em graus centígrados
	Corrente na fonte de carregamento	Corrente medida na fonte de carregamento em amperes
	Voltagem na fonte de carregamento	Voltagem medida na fonte de carregamento em Volts
	Tempo em segundos	Vetor de tempo por ciclo medido em segundos
Descarga	Voltagem medida	Voltagem do terminal da bateria medida em Volts
	Corrente medida	Corrente de saída da bateria medida em amperes
	Temperatura medida	Temperatura da bateria medida em graus centígrados
	Corrente na fonte de carregamento	Corrente medida no momento do carregamento em amperes
	Voltagem na fonte de carregamento	Voltagem medida no momento do carregamento em Volts
	Tempo em segundos	Vetor de tempo por ciclo medido em segundos
	Capacidade	A capacidade da bateria refere-se à quantidade de energia que a bateria consegue fornecer e armazenar após um ciclo de carga. A capacidade é expressa em amperes e é medida somente no final do ciclo de descarga.

Fonte: Chen et al. (2021)

O formato original dos conjuntos de dados é o formato *.mat*, por serem extraídos do MATLAB. Assim, para a sua leitura recorreu-se à biblioteca *scipy*.

De seguida, procurou-se aprofundar o entendimento dos conjuntos de dados, de maneira a desenhar uma estratégia para a extração de variáveis. Posto isto, os conjuntos de dados não processados, têm 11 colunas, nomeadamente: Ciclo, Tipo, Temperatura ambiente,

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Timestamp, Voltagem medida, Corrente medida, Temperatura medida, Corrente na fonte de carregamento, Voltagem na fonte de carregamento, Tempo em segundos e Capacidade.

Face ao facto de serem considerados os ciclos de carga e de descarga para a previsão, importa referir que uma vez que a capacidade é apenas observada no perfil de descarga, foi então considerado que a capacidade no ciclo n de carga será igual à capacidade no ciclo n de descarga.

Tendo em conta que o conjunto de dados contém dados de quatro baterias, de seguida são apresentadas as estatísticas descritivas de cada uma das baterias. Contudo, existe uma discrepância no número de registos entre os ciclos de carga e de descarga, o qual pode ser atribuída a fatores associados à forma como a recolha da informação foi feita durante experiência feita pela NASA.

2.2.1 Estatísticas descritivas da bateria B0005

A Tabela 2.2 contém as estatísticas descritivas das variáveis da bateria B0005.

Tabela 2.2 Estatísticas descritivas bateria B0005

Variáveis	Total	Média	Desvio Padrão	Mínimo	Máximo
Ciclo	591458	-	-	1	338
Carga	541173	-	-	1	170
Descarga	50285	-	-	1	168
Temperatura ambiente	591458	24	0	24	24
Timestamp	591458	-	-	02/04/2008 13:08:18	28/05/2008 11:09:42
Voltagem medida	591458	4,1	0,21	0	8,39
Corrente medida	591458	0,37	0,91	-4,48	1,53
Temperatura medida	591458	26,37	2,77	23,21	41,45
Corrente na fonte de carregamento	591458	0,63	0,74	-4,47	2
Voltagem na fonte de carregamento	591458	4,02	1,21	0	5
Tempo em segundos	591458	4763,86	3147,86	0	10807,33
Capacidade	591453	1,54	0,17	1,29	1,86

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Ciclo: A bateria B0005 tem no total 591458 registos de ciclos de carga e de descarga, dividindo-se em 170 ciclos de carga com 541173 registos e 168 ciclos de descarga com 50285 registos.

Voltagem medida: A variável voltagem medida tem uma média igual a 4,10V e um desvio padrão igual a 0,21V. Dado que varia entre 0V e 8,39V, verifica-se que a média se encontra aproximadamente a meio do intervalo.

Corrente medida: A média da corrente medida é 0,37A e o desvio padrão, 0,91A, que é significativamente maior do que a média, dando assim a indicação de que existe uma elevada variação nos valores da corrente medida.

Voltagem na fonte de carregamento: Esta variável tem uma média igual a 4,02V e uma variação entre 0V e 5V. À semelhança da voltagem medida, a variação entre os valores mínimos e máximo dá a entender a possível existência de *outliers*.

Corrente na fonte de carregamento: A variável corrente na fonte de carregamento varia entre os -4,47A a 2A. Os valores negativos dão a indicação da inversão da direção da corrente durante os ciclos de descarga.

Capacidade: O máximo da capacidade da bateria é 1,86Ah e o mínimo é 1,29Ah, indicando assim uma degradação ao longo dos ciclos de carga e de descarga.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

2.2.2 Estatísticas descritivas da bateria B0006

A Tabela 2.3 é representativa das estatísticas descritivas das variáveis da bateria B0006.

Tabela 2.3 Estatísticas descritivas bateria B0006

Variáveis	Total	Média	Desvio Padrão	Mínimo	Máximo
Ciclo	591458	-	-	1	338
Carga	541173	-	-	1	170
Descarga	50285	-	-	1	168
Temperatura ambiente	591458	24	0	24	24
Timestamp	591458	-	-	02/04/2008 13:08:18	28/05/2008 11:09:42
Voltagem medida	591458	4,1	0,22	0	8,08
Corrente medida	591458	0,36	0,88	-4,48	1,53
Temperatura medida	591458	26,43	2,78	22,98	42,01
Corrente na fonte de carregamento	591458	0,62	0,72	-4,48	2
Voltagem na fonte de carregamento	591458	4,24	0,73	-0,01	5
Tempo em segundos	591458	4763,86	3147,86	0	10807,33
Capacidade	591453	1,49	0,21	1,15	2,04

Ciclo: No total a bateria B0006 tem 591458 registos de ciclos de carga e de descarga. Os ciclos de carga têm um total 170 ciclos com 541173 registos, enquanto os ciclos de descarga têm um total de 168 ciclos com 50285 registos.

Voltagem medida: Existe uma similaridade na média e no desvio padrão da variável voltagem medida da bateria B0006 com os valores da bateria B0005. Em contrapartida, existe uma ligeira diferença em relação ao valor máximo da voltagem medida.

Corrente medida: O desvio padrão da variável corrente medida é significativamente maior do que a média, indicando assim, que existe uma alta variação nos valores da variável.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Voltagem na fonte de carregamento: Esta variável tem uma média e desvio padrão igual a 4,24V e 0,73V, respetivamente. A variação entre -0,01V e 5V, pode dar a indicação de *outliers*.

Corrente na fonte de carregamento: Existe uma similaridade nos valores da média, desvio padrão, mínimo e máximo da variável corrente na fonte de carregamento da bateria B0006 com os valores da bateria B0005. Os valores negativos dão a indicação da inversão da direção da corrente durante os ciclos de descarga.

Capacidade: O máximo da capacidade da bateria B0006 é superior à capacidade nominal de uma bateria, por 0,04Ah. A degradação da capacidade, ao longo dos ciclos de carga e de descarga, atingiu um mínimo de 1,15Ah.

As estatísticas descritivas das baterias B0007 e B0018 encontram-se no apêndice, na Tabela 0.1 e na Tabela 0.2, respetivamente.

3 PREPARAÇÃO DOS DADOS

A preparação dos dados é uma fase importante para o desenvolvimento de modelos de previsão, pois assegura a qualidade, consistência e relevância da informação utilizada para o treino e teste dos algoritmos. Neste capítulo, são descritos de forma detalhada os principais procedimentos realizados antes da modelação, incluindo a limpeza dos dados, a extração e criação de variáveis, a análise de correlação entre variáveis e, por fim, a seleção das variáveis mais relevantes para previsão do RUL.

3.1 Limpeza dos dados

A limpeza dos dados passou por:

- Remover ciclos inconsistentes;
- Remover ciclos com número de registos reduzidos.

Remoção de ciclos inconsistentes

Na fase da preparação dos dados, foram identificadas inconsistências na sequência dos ciclos de carga e de descarga. Em condições normais, os ciclos deveriam seguir uma alternância de carga – descarga – carga, o que reflete o comportamento normal na realização de testes em baterias.

Contudo, constatou-se que, em determinados pontos esta sequência não foi respeitada. Foram registados casos de ciclos consecutivos de carga (carga → carga → descarga) ou ciclos consecutivos de descarga (descarga → descarga → carga). Estes padrões são indicativos de possíveis interrupções ou ajustes manuais que ocorreram durante a execução dos testes.

A remoção dos ciclos inconsistentes passou pela análise do *timestamp* de modo a fazer uma seleção mais acertada dos ciclos que devem ser considerados. As Figura 3.1, Figura 3.2 e Figura 3.3 contém as informações sobre a inconsistência dos ciclos das baterias B0005, B0006 e B0007. Posto isto, foram eliminados os ciclos 12, 33 de carga e o ciclo 90 de descarga.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

# ciclo	# nova sequência de ciclos	tipo de ciclo	timestamp
21	11	charge	2008-04-04 06:58:00.296000
22	11	discharge	2008-04-04 09:57:19.765000
23	12	charge	2008-04-04 11:06:50.375000
24	13	charge	2008-04-04 15:05:59.906000
25	12	discharge	2008-04-04 17:56:27.609000

Figura 3.1 Ciclos inconsistentes de carga (12 e 13)

# ciclo	# nova sequência de ciclos	tipo de ciclo	timestamp
61	30	discharge	2008-04-21 02:15:02.921000
62	32	charge	2008-04-21 17:51:26.312000
63	33	charge	2008-04-22 14:15:41.187000
64	31	discharge	2008-04-22 15:33:49.875000

Figura 3.2 Ciclos inconsistentes de carga (32 e 33)

# ciclo	# nova sequência de ciclos	tipo de ciclo	timestamp
176	87	discharge	2008-05-07 16:59:29.937000
177	90	charge	2008-05-07 18:23:54.687000
178	88	discharge	2008-05-07 21:56:09.656000
179	91	charge	2008-05-07 23:20:34.515000
180	89	discharge	2008-05-08 02:53:49.937000
181	90	discharge	2008-05-09 12:25:07
182	92	charge	2008-05-09 17:26:51.968000

Figura 3.3 Ciclos inconsistentes de descarga (89 e 90)

Relativamente à bateria B0018, foram eliminados os ciclos 46 e 57 de carga e a informação sobre os ciclos inconsistentes são apresentadas na Figura 3.4 e Figura 3.5.

# ciclo	# nova sequência de ciclos	tipo de ciclo	timestamp
91	46	charge	2008-07-19 14:47:30.984000
92	47	charge	2008-07-29 17:16:52.453000
93	46	discharge	2008-07-29 18:34:27.281000
94	48	charge	2008-07-29 19:30:58.062000

Figura 3.4 Ciclos inconsistentes de carga (46 e 47)

# ciclo	# nova sequência de ciclos	tipo de ciclo	timestamp
111	55	discharge	2008-07-31 10:10:14.359000
112	57	charge	2008-07-31 15:20:13.937000
113	58	charge	2008-08-01 17:14:25.234000
114	56	discharge	2008-08-01 17:46:33.109000

Figura 3.5 Ciclos inconsistentes de carga (57 e 58)

Remoção do ciclo incompleto

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Para as baterias B0005, B0006 e B0007, foi removido o último ciclo de carga, uma vez que o mesmo é um ciclo incompleto por apresentar um número reduzido de registos. A Figura 3.6 contém o número total de registos do último ciclo de carga.

```
<class 'pandas.core.frame.DataFrame'>
Index: 5 entries, 591453 to 591457
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   cycle        5 non-null      int64
1   timestamp    5 non-null      datetime64[ns]
2   Time         5 non-null      float64
dtypes: datetime64[ns](1), float64(1), int64(1)
memory usage: 160.0 bytes
```

Figura 3.6 Total número de registos do ciclo 170 de carga

Após a limpeza dos dados, foi adicionada uma nova variável que no fundo representa a nova sequência de ciclos de 1 a 167 para os ciclos de carga e descarga das baterias B0005, B0006 e B0007. Para a bateria B0018, o processo também foi igual, mas tendo esta bateria um número reduzido de ciclos, a sequência dos ciclos passou a ser de 1 a 132 ciclos de carga e descarga.

3.2 Extração de variáveis

Após a limpeza dos dados, procedeu-se à extração das variáveis relevantes para o estudo. Considerando que o objetivo principal é a previsão do RUL, optou-se por trabalhar com informação agregada por ciclo. Esta abordagem permite representar de forma mais consistente o comportamento de cada bateria ao longo do tempo, reduzindo a variabilidade associada aos registos individuais de cada ciclo de carga e descarga.

A presente dissertação teve como inspiração a extração de variáveis feita por Chen et al. (2021), em que foram extraídas 20 variáveis para a realização de previsão. No entanto, foram adicionados mais nove variáveis, nomeadamente: tempo total por ciclo de carga e o tempo total por ciclo de descarga, o nome da bateria, ciclo carga, ciclo descarga, capacidade, SoH e RUL, perfazendo assim um total de 29 variáveis. A Tabela 3.1 contém a listagem de todas as variáveis extraídas e criadas que serão descritas nesta secção.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

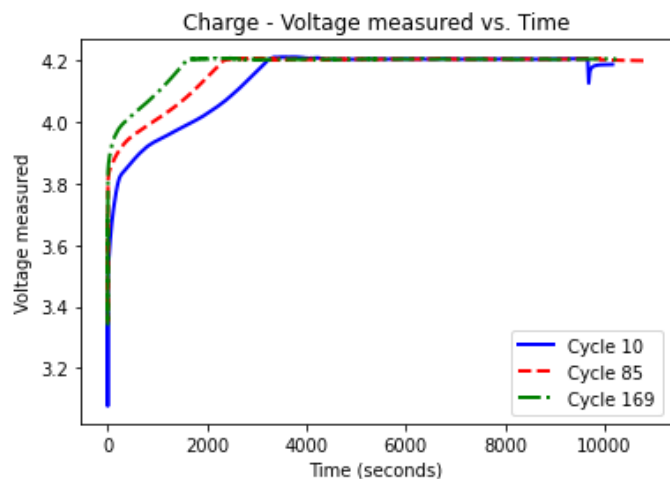
Tabela 3.1 Variáveis de carga e descarga.

Perfil	Código da variável	Nome
Carga e Descarga	<i>Battery_id</i>	Nome da bateria
Carga e Descarga	Ciclo	Ciclo de carga e descarga
Carga	<i>Type_C</i>	Perfil Carga
Carga	V1	Voltagem medida
Carga	V2	O primeiro instante em que se regista a voltagem máxima medida
Carga	V3	Corrente medida
Carga	V4	Instante em que se regista o início da diminuição da corrente medida
Carga	V5	Temperatura medida
Carga	V6	Instante em que se regista a temperatura máxima medida
Carga	V7	Corrente na fonte de carregamento
Carga	V8	Instante em que se regista a voltagem máxima da fonte de carregamento
Carga	V9	Voltagem na fonte de carregamento
Carga	V10	Instante em que se regista o início da redução da corrente na fonte do carregamento
Carga	V11	Tempo total por ciclo de carga
Descarga	<i>Type_D</i>	Perfil Descarga
Descarga	V12	Voltagem medida
Descarga	V13	O primeiro instante em que se regista o valor mínimo da voltagem medida
Descarga	V14	Corrente medida
Descarga	V15	Instante em que se regista a subida da corrente medida
Descarga	V16	Temperatura medida
Descarga	V17	Instante em que se regista a temperatura máxima medida
Descarga	V18	Corrente na fonte de carregamento
Descarga	V19	Instante em que se regista o ponto em que se regista a descida da corrente da fonte de carregamento
Descarga	V20	Voltagem na fonte de carregamento
Descarga	V21	Instante em que se regista o valor da voltagem mínima na fonte de carregamento
Descarga	V22	Tempo total por ciclo de descarga
Carga e Descarga	Capacidade	Capacidade da bateria ao longo dos ciclos
Carga e Descarga	SoH	Índice de estado de saúde da bateria
Carga e Descarga	RUL	Vida útil restante da bateria

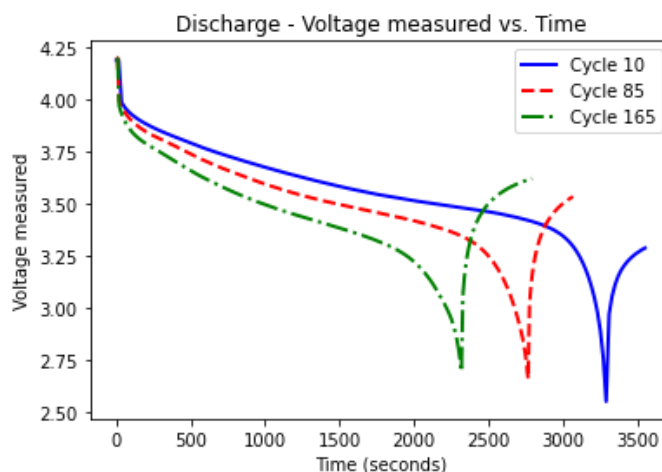
Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

De modo a não tornar apresentação visual das variáveis muito repetitiva, serão apenas ilustradas, graficamente, as variáveis da bateria B0005.

Voltagem medida



(a)



(b)

Figura 3.7 Voltagem medida ao longo de três ciclos de carga (a) e ao longo de três ciclos de descarga (b)

A Figura 3.7 (a) ilustra, para três ciclos de carga, a relação entre o tempo e a voltagem medida. Observa-se que, o tempo para carregar a bateria até a voltagem máxima, vai diminuindo à medida que, o número de ciclos aumenta, portanto, quanto mais vezes a bateria é carregada, menor é o tempo necessário para atingir a voltagem máxima no ciclo de carga. Esta relação aponta para potenciais preditores, tendo assim sido extraídas duas variáveis, nomeadamente o valor máximo da voltagem (variável V1 da Tabela 3.1) e o primeiro instante em que se regista o valor máximo de voltagem (variável V2 da Tabela 3.1).

A Figura 3.7 (b) representa, para três ciclos de descarga, a relação entre o tempo e a voltagem medida. Observa-se quanto maior for o número de ciclos de descarga menor será o tempo necessário para atingir a voltagem mínima no ciclo de descarga. De forma análoga são extraídas as variáveis, valor mínimo da voltagem (variável V12 da Tabela 3.1) e o primeiro instante em que se regista o valor mínimo de voltagem (variável V13 da Tabela 3.1).

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Corrente medida

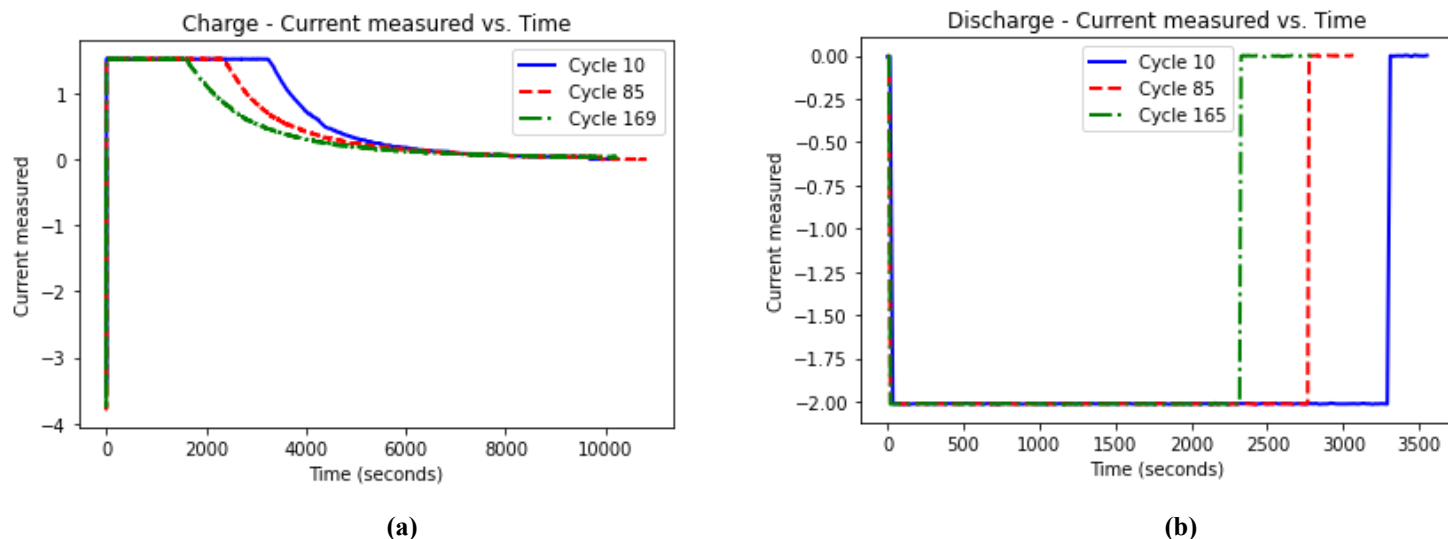


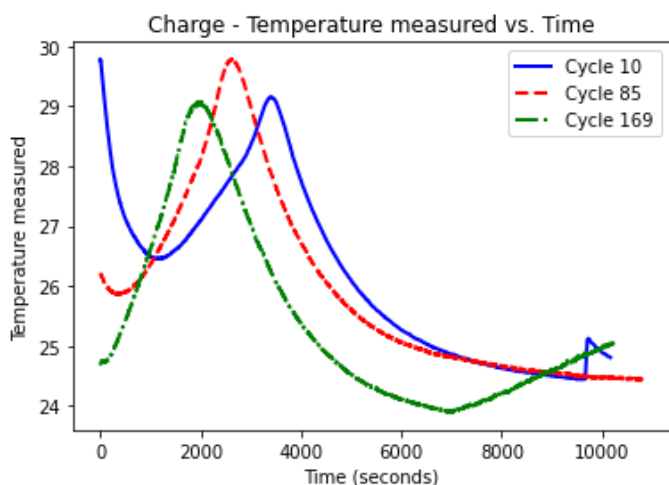
Figura 3.8 Corrente medida de três ciclos de carga (a) e três ciclos de descarga (b)

A Figura 3.8 (a) exhibe, para três ciclos de carga, a relação entre o tempo e a corrente medida. Verifica-se que o ponto em que se inicia a redução da corrente tende a ocorrer mais cedo à medida que o número de ciclos aumenta. De forma análoga são extraídas duas variáveis, o ponto de quebra, imediatamente antes do início da diminuição da corrente medida (variável V3 da Tabela 3.1) e o instante em que se indica a redução da corrente (variável V4 da Tabela 3.1).

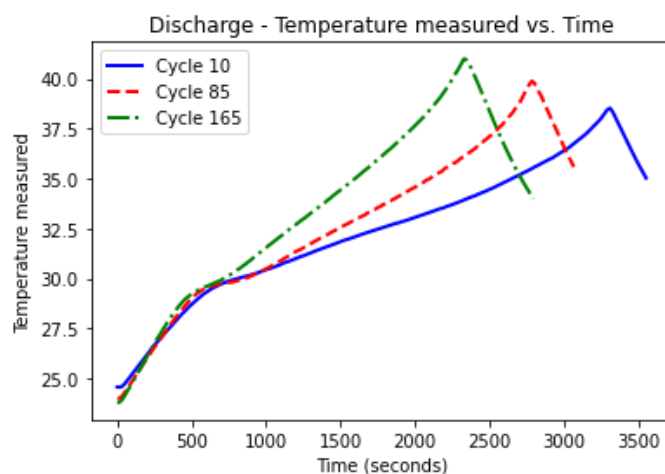
A Figura 3.8 (b) ilustra, para os ciclos de descarga, a relação entre o tempo e a corrente medida. Verifica-se que o tempo decorrido desde o início da descarga até à recuperação da corrente diminui progressivamente com o aumento dos ciclos. Desta relação foram extraídas duas variáveis adicionais: o ponto em que ocorre a subida da corrente medida (variável V14 da Tabela 3.1) e o instante em que se regista a subida (variável V15 da Tabela 3.1).

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Temperatura medida



(a)



(b)

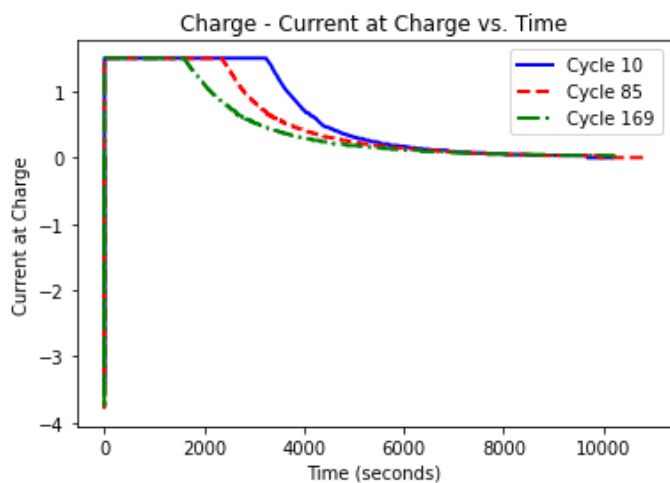
Figura 3.9 Temperatura medida de três ciclos de carga (a) e três ciclos de descarga (b)

Na Figura 3.9 (a) observa-se, para os ciclos de carga, a relação entre o tempo e a temperatura medida. Constatase que o tempo necessário para atingir a temperatura máxima tende a diminuir com o aumento dos ciclos. Em alguns casos, a temperatura máxima é atingida logo no início do processo; nesses casos, foi considerado o segundo pico máximo, isto é, o valor registado após o marco dos 1000 segundos, garantindo uma leitura mais representativa da evolução térmica. Deste modo, foram extraídas duas variáveis: a temperatura máxima medida após os primeiros 1000 segundos (variável V5 da Tabela 3.1) e o instante, em segundos, até à sua ocorrência (variável V6 da Tabela 3.1).

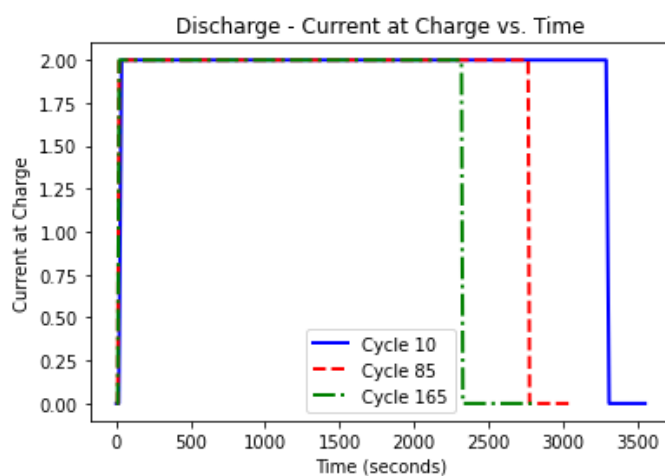
A Figura 3.9 (b) evidencia, novamente a relação entre a temperatura medida e o tempo, para os ciclos de descarga, reforçando a tendência de redução do tempo necessário para alcançar a temperatura máxima à medida que o número de ciclos aumenta. De forma análoga são extraídas duas variáveis, o ponto de início da diminuição da corrente medida (variável V16 da Tabela 3.1) e o instante em que se dá o registo (variável V17 da Tabela 3.1).

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Corrente na fonte de carregamento



(a)



(b)

Figura 3.10 Corrente na fonte de carregamento de três ciclos de carga (a) e três ciclos de descarga (b)

A Figura 3.10 (a) apresenta, para os ciclos de carga, a relação entre o tempo e a corrente na fonte de carregamento. Observa-se uma redução progressiva do tempo necessário até ao início da diminuição da corrente. Foram, portanto, extraídas duas variáveis: o ponto de início da redução da corrente na fonte de carregamento (variável V7 da Tabela 3.1) e o instante temporal correspondente (variável V8 da Tabela 3.1).

Durante a descarga, conforme ilustrado na Figura 3.10 (b), o tempo necessário para atingir a corrente mínima na fonte de carregamento também apresenta uma redução ao longo dos ciclos. Desta relação resultaram duas variáveis adicionais: o ponto em que se regista a descida da corrente da fonte de carregamento (variável V18 da Tabela 3.1) e o instante, em segundos, até à sua ocorrência (variável V19 da Tabela 3.1).

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Voltagem na fonte do carregamento

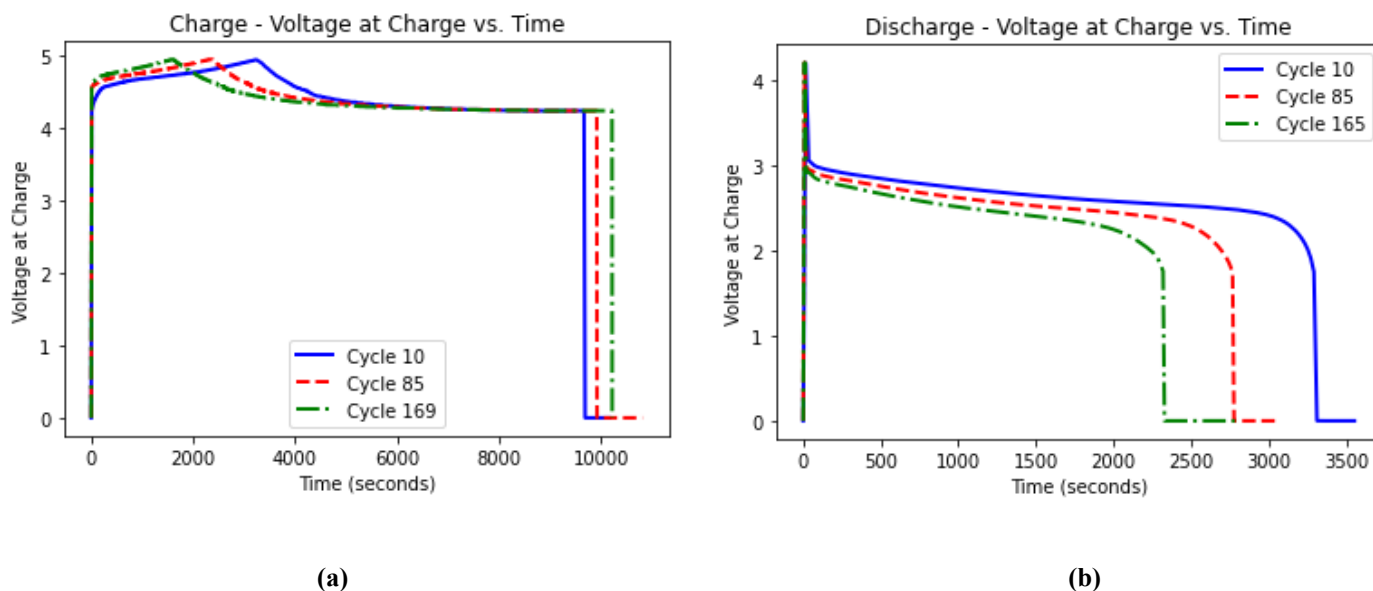


Figura 3.11 Voltagem máxima na fonte de três ciclos de carga (a) e três ciclos de descarga (b)

A Figura 3.11 (a) evidencia, para os ciclos de carga, a relação entre o tempo e a voltagem na fonte de carregamento. Verifica-se que o tempo necessário para atingir a voltagem máxima na fonte de carregamento diminui progressivamente à medida que o número de ciclos aumenta. Deste modo, foram extraídas a voltagem máxima da fonte de carregamento (variável V9 da Tabela 3.1) e o instante, em segundos, correspondente (variável V10 da Tabela 3.1).

Na Figura 3.11 (b) é apresentada a relação entre o tempo e a voltagem na fonte de carregamento para os ciclos de descarga. Nota-se que a voltagem mínima tende a ocorrer cada vez mais cedo. Contudo, dado que em alguns ciclos o valor mínimo se regista logo no início, optou-se por considerar o segundo ponto mínimo, isto é, o valor obtido após os 1000 segundos, de forma a garantir uma análise mais consistente. De forma análoga são extraídas duas variáveis, a voltagem mínima na fonte de carregamento (variável V20 da Tabela 3.1) e o tempo necessário para a atingir (variável V21 da Tabela 3.1).

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Tempo total

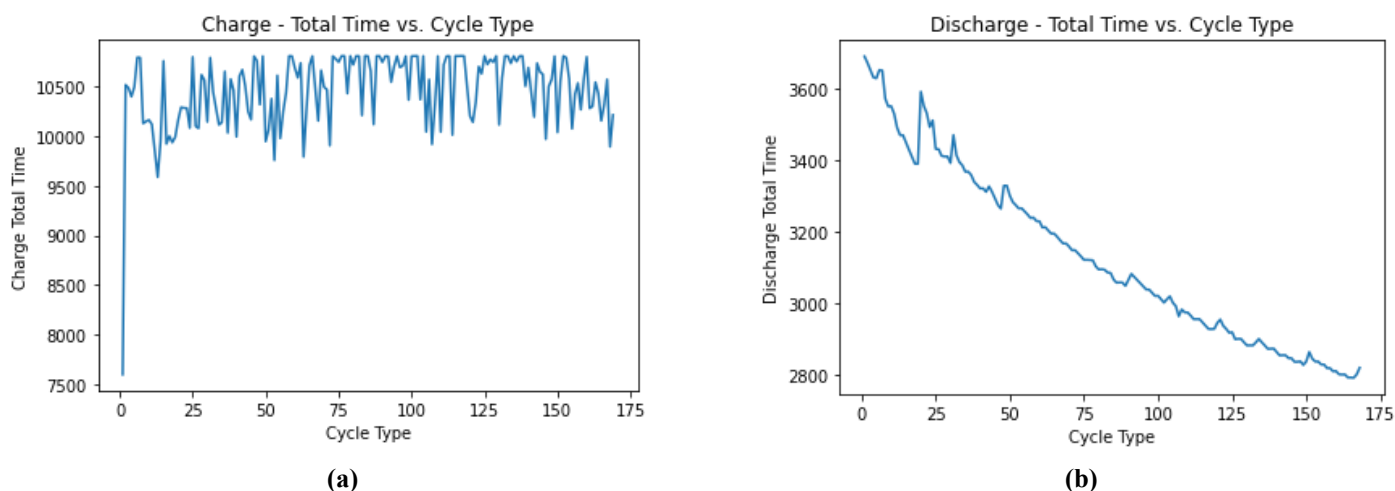


Figura 3.12 Tempo total de cada ciclo de carga (a) e de descarga (b)

O tempo total de cada ciclo de carga e de descarga corresponde ao intervalo que decorre entre o início o fim de cada ciclo. Ou seja, o tempo total foi obtido a partir da diferença entre os instantes registados no início e no fim de cada ciclo, o que permite quantificar a duração efetiva de cada ciclo de carga e de descarga. A Figura 3.12 (a) e (b) representam o tempo total do ciclo de carga (variável V11 da Tabela 3.1) e o tempo total do ciclo de descarga (variável V22 da Tabela 3.1).

SoH

Segundo Chen et al. (2021) a fórmula para calcular o SoH é:

$$SoH = \frac{\text{Capacidade real}}{\text{Capacidade nominal}} \quad (3.1)$$

onde a capacidade atual é capacidade no ciclo n e capacidade nominal usada é 2 Ah, ou seja, a capacidade da bateria quando nova.

RUL

A fórmula que foi utilizada para calcular o RUL, segundo Chen et al. (2021), é:

$$RUL = Ciclo_n - Ciclo_{atual} \quad (3.2)$$

onde o $Ciclo_n$ corresponde ao ciclo onde as baterias atingem o ponto de EoL, e o $Ciclo_{atual}$, tal como o nome sugere, corresponde ao número do ciclo atual.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Informações suplementares

Para a extração das variáveis V3, V4, V7, V8, V14, V15, V20 e V21 recorreu-se à biblioteca *Ruptures* em *Python* introduzida por (Truong et al., 2020). *Ruptures* é uma biblioteca que permite a deteção do ponto de quebra, ou seja, do ponto onde surge uma mudança num modelo de series temporais ou sinais. Posto isto, através da utilização de *Ruptures*, para cada ciclo de carga e descarga, foram extraídos os pontos de quebra e o respetivo instante em que se dá o registo.

Por fim, apesar de cada uma das baterias ter ciclos de carga e de descarga após o ponto onde é atingido o EoL, os mesmos serão desconsiderados uma vez se que pretende utilizar os ciclos da degradação das baterias até a perda da capacidade inicial, ou seja, até a perda de 30% da capacidade.

Estatísticas descritivas das variáveis criadas

- As estatísticas descritivas da Tabela 3.2 são referentes à concatenação dos valores das variáveis criadas das baterias B0005, B0006 e B0007 que serão utilizadas para o conjunto de treino.

A partir destas estatísticas descritivas, se pode constatar que:

- O número de ciclos varia de 1 a 163, correspondendo ao número de ciclos de carga e de descarga.
- A capacidade diminui de 2,04Ah para 1,4Ah, indicando assim a degradação da bateria ao longo dos ciclos.
- Para os ciclos de carga, os registos das variáveis V1 e V9 permanecem dentro de intervalos estreitos. À semelhança acontece para a variável V3 com um intervalo de 1,42A a 1,52A.
- Existe uma grande variação na variável V16 do ciclo de descarga que tem uma média de temperatura = 37,31°C e um desvio padrão = 4,51°C, o que indica que existem flutuações térmicas durante os ciclos de descarga.
- As variáveis que representam os instantes em segundo, como por exemplo, a variável V8 do ciclo de carga e V17 do ciclo de descarga, apresentam grandes

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

desvio padrão, o que sugere que existe uma variação significativa nas durações dos ciclos de carga e de descarga.

Contudo, é possível notar em média que, os ciclos de carga tendem a ser mais longos do que os ciclos de descarga.

Tabela 3.2 Estatísticas descritivas das variáveis criadas

Código da variável	Total	Média	Desvio Padrão	Mínimo	Máximo
Ciclo	395	-	-	1	163
V1	395	4,21	0	4,2	4,22
V2	395	5358,22	2523,91	1421,48	9538,16
V3	395	1,49	0,02	1,42	1,52
V4	395	2652,41	554,74	717,52	3608,81
V5	395	29,21	0,59	27,26	31,19
V6	395	2926,82	483,38	1006,09	3894,33
V7	395	1,45	0,11	0,85	1,51
V8	395	2690,87	582,12	717,52	4140,62
V9	395	4,94	0,01	4,92	4,96
V10	395	2612,98	550,88	653,83	3604,92
V11	395	10452,16	403,7	7597,88	10807,33
V12	395	2,34	0,25	1,74	2,7
V13	395	3023,08	306,12	2512,7	3690,23
V14	395	-1,66	0,76	-2,02	0
V15	395	2795,91	631,76	141,25	3471,02
V16	395	37,31	4,51	28,67	42,33
V17	395	2473,14	933,33	1000,48	3507,33
V18	395	1,19	0,98	0	2
V19	395	2862,21	533,82	362,58	3570,45
V20	395	0,19	0,48	0	1,72
V21	395	3034,05	306,46	2522,33	3690,23
V22	395	3201,37	228,21	2793,39	3690,23
capacidade	395	1,66	0,17	1,4	2,04
SoH	395	0,83	0,08	0,7	1,02
RUL	395	-	-	1	163

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

3.3 Matriz de correlação

Antes de realizar a seleção de variáveis, foi criada uma matriz de correlação para examinar as relações entre as variáveis no conjunto de dados concatenado das três baterias. A combinação dos dados permitiu uma visão mais ampla das dependências e interações entre as variáveis em diferentes condições da bateria.

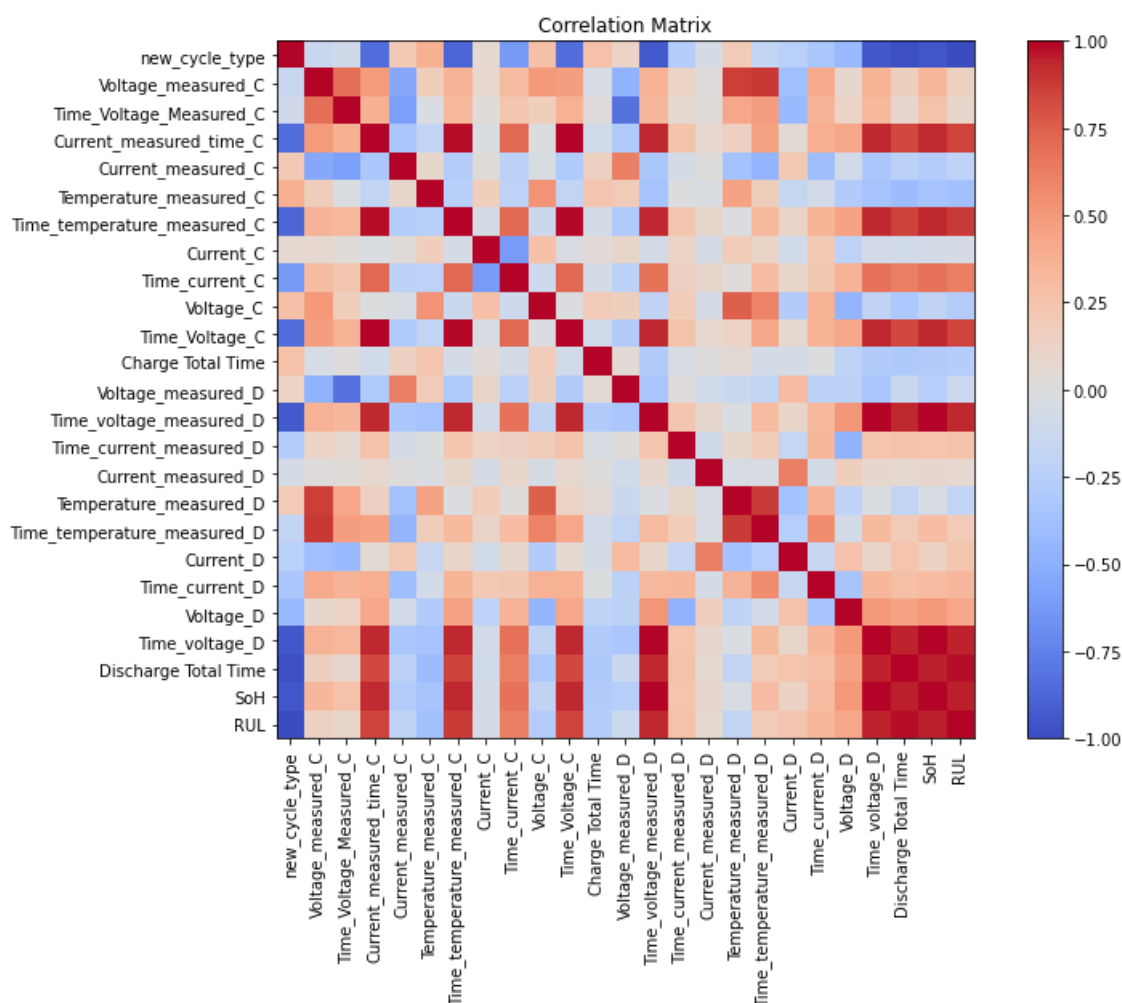


Figura 3.13 Matriz de correlação

A Figura 3.13 exibe o coeficiente de correlação linear entre todas as variáveis no conjunto de dados combinado. Cada célula representa o coeficiente de correlação entre duas variáveis, com valores próximos a +1 indicando fortes correlações positivas e valores próximos a -1 indicando fortes correlações negativas. A visualização do mapa de calor

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

destaca essas relações, onde os tons de vermelho mais carregados representam correlações positivas e os tons de azul mais carregados indicam correlações negativas.

A análise da matriz de correlação ajudou a identificar as variáveis altamente correlacionadas com o RUL, orientando a seleção das variáveis mais relevantes para a previsão.

3.4 Seleção de variáveis

Para melhorar a precisão dos modelos de previsão, recorreu-se ao método *SelectKBest* da biblioteca *scikit-learn* para escolher as variáveis mais relevantes. O *SelectKBest* atribui uma pontuação a cada variável independente com base na relação com a variável dependente, em seguida, mantém as k variáveis com as pontuações mais altas. O valor de k representa o número de variáveis com melhor desempenho que são selecionadas para a previsão (scikit-learn developers, s.d.).

Para o presente estudo, foi utilizada a função de pontuação $f_{\text{regression}}$. Esta função mede como cada variável independente está linearmente relacionada à variável dependente, através do cálculo de uma estatística F e um valor p . As variáveis com pontuações F mais altas são mais úteis para a previsão, pois têm uma forte correlação com a variável dependente (scikit-learn developers, s.d.).

A partir da análise, verificou-se que nove variáveis estavam fortemente correlacionadas com o RUL. No entanto, apenas cinco foram selecionadas para a realização da previsão, nomeadamente:

- Ciclos;
- Instante em que se regista o valor mínimo de voltagem nos ciclos de descarga;
- Instante em que se regista o valor mínimo da voltagem medida nos ciclos de descarga;
- O tempo total em segundos de cada ciclo de descarga;
- SoH.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Esta decisão baseou-se na natureza do conjunto de dados experimentais, em que muitas variáveis tendem a estar inter-relacionadas, indicando fortes correlações entre elas. A seleção apenas das cinco principais variáveis ajudou a reduzir a redundância e a multicolinearidade, garantindo que os modelos aprendessem com as variáveis mais informativas e independentes.

4 MODELAÇÃO

Neste capítulo serão apresentados o modelo e os algoritmos utilizados para a realização da previsão, as ferramentas de otimização de hiperparâmetros e as iterações feitas.

Para a realização da previsão do RUL da bateria, serão utilizados três modelos de previsão, nomeadamente: RL, que será usado como modelo *baseline* (i.e., um modelo naïve que é usado para perceber o valor acrescentado dos modelos mais complexos), XGBoost que é o algoritmo principal da presente dissertação e por fim um algoritmo comparativo que será a RF. Sendo que estes últimos são algoritmos baseados em árvore, o objetivo passará pela verificação da existência de alguma diferença significativa entre eles.

O treino destes modelos passará pela divisão dos dados para treino e teste, em que serão utilizados, os dados das baterias B0005, B0006 e B0007 para o conjunto de treino e a bateria B0018 para o conjunto de teste.

Por fim, o pipeline da modelação encontra-se no APÊNDICE 2. Pipeline da modelação.

4.1 Modelos

4.1.1 Modelo *baseline*

A RL é um modelo clássico de ML utilizado para e investigar a relação linear existente entre as variáveis independentes e a variável dependente, ou a que se pretende prever (Arena et al., 2022). O modelo de RL tem a seguinte fórmula:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N + \varepsilon \quad (4.1)$$

onde:

- y é a variável que se pretende prever;
- $x_1, x_2, \dots, x_N$ são as variáveis independentes;
- $\beta_0, \beta_1, \beta_2, \dots, \beta_N$ são os parâmetros do modelo;
- ε – é o termo erro ou termo de perturbação aleatório.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

4.1.2 RF

O RF é um algoritmo de ML que combina várias DT de modo a melhorar a precisão da previsão e reduzir o sobreajustamento. O algoritmo segue o conceito de *Bootstrap Aggregation*, ou seja, cada uma é concebida com recurso a um subconjunto aleatório distinto de preditores e um subconjunto de dados de treino aleatório. A previsão final do RF é a média de todos os resultados das DT (Jafari et al., 2024; Wang et al., 2023).

O algoritmo RF introduz:

- **Bagging:** Cada DT é treinada em um subconjunto aleatório de dados.
- **Seleção aleatória de variáveis:** Um subconjunto aleatório de variáveis é considerado em cada divisão de nó.
- **Ensembling:** Cada DT criada pelo algoritmo é um “modelo fraco”, contudo, a agregação das DT reduz a variância.

Vantagens do RF segundo Wang et al. (2023):

- **Modelação não linear:** O RF pode capturar dependências não lineares entre variáveis independentes e a variável dependente.
- **Robustez/redução de sobreajustamento:** Em comparação com uma única DT a agregação de DT ajuda a reduzir sobreajustamento.
- **Capacidade de interpretação relativa:** A possibilidade de calcular medidas de importância das variáveis de modo a avaliar quais é que têm mais influência.
- **Simples de ajustar os hiperparâmetros:** A RF tem um número reduzido de parâmetros em comparação aos modelos de *boosting*; os valores pré-definidos, de um modo geral, têm um bom desempenho.

Com isto, o algoritmo é mais robusto em comparação com uma única DT e é menos sujeito ao sobreajustamento.

4.1.3 XGBoost

De acordo com Jafari et al. (2023), o XGBoost é conhecido pelas suas capacidades de previsão excecionais e é um modelo de conjunto baseado em árvores de elevado

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

desempenho. O XGBoost é um algoritmo de ML avançado do método de *gradient boosting* que combina, de forma sequencial, múltiplas DT para criar um modelo mais forte e robusto. Cada modelo seguinte tenta corrigir os erros realizados pelos modelos anteriores, adicionando peso aos dados mal classificados (Warni et al., 2024).

Segundo Abiodun et al. (2024), Chen e Guestrin (2016) e Deng e Lumley (2023), o algoritmo XGBoost introduz:

1. **Regularização L1 e L2:** O algoritmo incorpora termos de regularização na sua função objetivo, de uma forma semelhante à regressão *Ridge* (L2) e à regressão *Lasso* (L1), o que permite controlar a complexidade do algoritmo e prevenir o sobreajustamento.
2. **Eficiência computacional e paralelização:** Durante a construção das árvores o algoritmo implementa paralelização de modo a reduzir o tempo de treino, o que permite lidar com grandes volumes de dados.
3. **Tratamento de dados dispersos e de valores em falta:** O algoritmo lida, de forma automática, com dados em falta ou valores dispersos, identificando o caminho de divisão mais apropriado durante o treino.
4. **Aproximação na construção das árvores:** De modo a reduzir a complexidade computacional sem comprometer o seu desempenho, o XGBoost utiliza métodos aproximados para determinar os pontos de divisão ideais.

O XGBoost apresenta as seguintes vantagens, conforme destacado por Chen e Guestrin (2016), Kiriakidou et al. (2024), Xiao et al. (2024) e Xu et al. (2025):

- **Elevada precisão da previsão:** Combina *boosting* e regularização para alcançar uma elevada precisão de previsão, robustez e capacidade de generalização, minimizando o sobreajustamento.
- **Eficiência computacional:** A sua implementação otimizada e paralelizada garante um desempenho significativamente mais rápido.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- **Lida com valores em falta:** Lida automaticamente com valores em falta, determinando o melhor caminho, o que reduz a necessidade de um pré-processamento intensivo dos dados.
- **Facilidade de interpretação de variáveis:** Permite uma fácil interpretação da importância das variáveis, clarificando o impacto de cada uma nas previsões do modelo.
- **Escalabilidade:** Oferece um elevado nível de escalabilidade para grandes volumes de dados e é facilmente integrável em *pipelines* de monitorização contínua.

4.2 Hiperparâmetros

Para o treino de um modelo de ML é necessária uma seleção de hiperparâmetros (i.e., parâmetros definidos antes do treino e que são responsáveis pelo controlo do comportamento e pela estrutura de um modelo de ML²) de modo a melhorar o desempenho.

A otimização dos algoritmos selecionados para a previsão do RUL passou pela utilização de duas técnicas, nomeadamente: *GridSearchCV* e *Optuna*, com o propósito de comparar a eficiência e o desempenho de cada técnica.

GridSearchCV

O *GridSearchCV* faz uma pesquisa exaustiva sobre um conjunto predefinido de valores possíveis para cada hiperparâmetro. Esta técnica tem a função de avaliar todas as combinações de parâmetros com recurso à validação cruzada, selecionando assim a combinação que apresenta o melhor desempenho segundo uma métrica de avaliação como por exemplo a raiz quadrada do erro médio. Importa referir que apesar desta técnica ser robusta e de fácil interpretação, existe um elevado custo computacional particularmente quando se trata de pesquisas muito extensas (scikit-learn developers, s.d.).

² Fonte: <https://aws.amazon.com/what-is/hyperparameter-tuning/>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Optuna

O *Optuna* é uma biblioteca moderna que adota uma abordagem, facilmente adaptável e inteligente, de otimização automática de hiperparâmetros. Esta técnica utiliza técnicas de amostragem sequencial e algoritmos de otimização bayesiana de modo a explorar eficientemente o espaço de hiperparâmetros. Através de iterações, o *Optuna* ajusta a pesquisa tendo como base em resultados anteriores, mantendo o foco em combinações que possam gerar melhores resultados. Posto isto, o *Optuna* destaca-se como uma ferramenta eficiente em termos do tempo de processamento e é adequada a problemas complexos e de elevada dimensionalidade (Akiba et al., 2019).

4.2.1 Hiperparâmetros do algoritmo RF

Para o algoritmo RF foram selecionados quatro hiperparâmetros (scikit-learn developers, s.d.) que são apresentados de seguida:

- ***n_estimators***: corresponde ao número de DT no algoritmo RF. Os valores mais elevados tendem a reduzir a variância, em contrapartida, aumentam o tempo de processamento.
- ***max_depth***: representa a profundidade máxima permitida para cada DT. Este hiperparâmetro controla a complexidade do algoritmo, prevenindo assim, o sobreajustamento.
- ***max_features***: corresponde ao número de variáveis consideradas em cada divisão. A influência a diversidade entre as DT que este hiperparâmetro tem contribui para a estabilidade do modelo.
- ***min_samples_split***: representa o número mínimo de amostras necessário para a divisão de um nó. Os valores mais elevados resultam em algoritmos mais generalizáveis e simples.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

A Tabela 4.1 contém a listagem dos hiperparâmetros, os valores pré-definidos e os intervalos de valores admissíveis para treinar o algoritmo RF.

Tabela 4.1 Hiperparâmetros do algoritmo RF

Hiperparâmetros	Valores pré-definidos	Intervalo de valores admissíveis
<i>Maximum_depth</i>	<i>None</i>	$[0, +\infty[$
<i>n_estimators</i>	100	$[100, 500]$
<i>max_features</i>	<i>Sqrt</i>	<i>Sqrt/log2/None</i>
<i>Min_sample_split</i>	2	$[1, 10]$

4.2.2 Hiperparâmetros do algoritmo XGBoost

Para o algoritmo XGBoost foram selecionados quatro hiperparâmetros (XGBoost Developers, s.d.) que são apresentados de seguida:

- ***learning_rate***: é responsável por regular o impacto de cada DT adicionada. Importa referir que, taxas mais baixas produzem um modelo mais estável, mas exigem mais iterações.
- ***max_depth***: corresponde à profundidade máxima das DT. Este hiperparâmetro define a complexidade do algoritmo, devendo ser ajusta de maneira cautelosa para evitar sobreajustamento.
- ***n_estimators***: Corresponde ao número de árvores no algoritmo XGBoost. À semelhança deste hiperparâmetro para o algoritmo RF, valores elevados tendem a reduzir a variância, mas aumentam o tempo de processamento.
- ***subsample***: representa a fração de amostras utilizadas em cada iteração, o que introduz variabilidade e melhora a generalização.
- ***objective***: é função de perda utilizada para medir o erro. Tratando-se de regressão, foi adotada a função *reg:squarederror*.
- ***booster***: indica o tipo de algoritmo de *boosting* utilizado, sendo *gbtree* o mais comum para a realização de tarefas de regressão.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

A Tabela 4.2 contém a listagem dos hiperparâmetros, os valores pré-definidos e os intervalos de valores admissíveis para treinar o algoritmo o XGBoost.

Tabela 4.2 Hiperparâmetros do algoritmo XGBoost

Hiperparâmetros	Valores pré-definidos	Intervalo de valores admissíveis
<i>Learning_rate (eta)</i>	0.3	[0,1]
<i>Maximum_depth</i>	6	[0, + ∞ [
<i>n_estimators</i>	100	[100,1000]
<i>Subsample</i>	1	[0,1]
<i>Objective</i>	<i>squarederror</i>	-
<i>Boost</i>	<i>Gbtree</i>	-

4.3 Iterações

A modelação passou por três iterações:

- Iteração 1 – Treinar o modelo *baseline* e os algoritmos RF e XGBoost com valores pré-definidos dos hiperparâmetros.
- Iteração 2 – Otimização dos hiperparâmetros e treinar os algoritmos RF e XGBoost com os hiperparâmetros otimizados.
- Iteração 3 – Normalização da variável RUL.

4.3.1 Iteração 1

O ponto de partida da modelação passou pela realização da previsão do RUL da bateria B0018 considerando apenas os hiperparâmetros com valores pré-definidos para os algoritmos RF e XGBoost.

4.3.2 Iteração 2

A segunda iteração passou pela otimização dos hiperparâmetros dos modelos RF e XGBoost, de modo a perceber se efetivamente existe uma notável melhoria em comparação com o desempenho que os modelos tiveram na primeira iteração. Para este propósito, foram treinados os algoritmos otimizados e feita a comparação dos resultados obtidos com os da primeira iteração.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

4.3.3 Iteração 3

Com o objetivo de melhorar a previsão, e tendo em consideração as diferenças que existem em termos do RUL das baterias, propôs-se uma estratégia que visa à normalização (min-max) da variável RUL do conjunto de dados de treino.

5 RESULTADOS

Neste capítulo serão primeiro apresentados as métricas de avaliação selecionadas e os resultados da previsão do RUL. A análise dos resultados, tal como mencionado na última subsecção do capítulo 4, está organizada em três iterações, em que são comparados os resultados obtidos em cada uma delas.

5.1 Métricas de avaliação

Para a avaliação do desempenho do modelo e algoritmos utilizados para fazer a previsão do RUL foram selecionadas três métricas de avaliação, nomeadamente: raiz quadrada do erro médio (RMSE, do inglês *Root Mean Squared Error*), erro percentual absoluto médio (MAPE, do inglês *Mean Absolute Percentage Error*) e erro absoluto médio (MAE, do inglês *Mean Absolute Error*).

Para o cálculo das métricas mencionadas acima, de seguida são apresentadas as descrições das variáveis que compõem as respetivas fórmulas:

- n corresponde ao número de ciclos;
- RUL_n é o valor real;
- $\widehat{RUL}_n$ é o valor previsto.

RMSE:

O RMSE mede a distância média entre o valor real e o valor previsto.

O RMSE é calculado através da seguinte fórmula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{n=1}^n (RUL_n - \widehat{RUL}_n)^2} \quad (5.1)$$

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

MAE:

Esta métrica calcula a magnitude média dos erros de previsão sem considerar a direção dos erros.

O MAE é calculado através da seguinte fórmula:

$$MAE = \frac{1}{n} \sum_{n=1}^n |RUL_n - \widehat{RUL}_n| \quad (5.2)$$

MAPE:

O MAPE expressa os erros de previsão como percentagens dos valores reais, tornando-o adequado para comparação entre diferentes conjuntos de dados.

O MAPE é calculado através da seguinte fórmula:

$$MAPE = \frac{1}{n} \sum_{n=1}^n \frac{|RUL_n - \widehat{RUL}_n|}{RUL_n} \quad (5.3)$$

5.2 Iteração 1

Na primeira iteração o modelo *baseline* e os algoritmos RF e XGBoost foram treinados individualmente sobre o mesmo conjunto de dados de treino para prever o RUL da bateria B0018. Importante referir que os resultados das métricas de desempenho que serão apresentados são apenas para o conjunto de teste.

5.2.1 Modelo *baseline*

O modelo *baseline* apresentou um RMSE de 29,25, um MAE de 28,64 e um MAPE de 189,13%, indicando uma discrepância significativa entre os valores do RUL real e os valores do RUL previsto.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

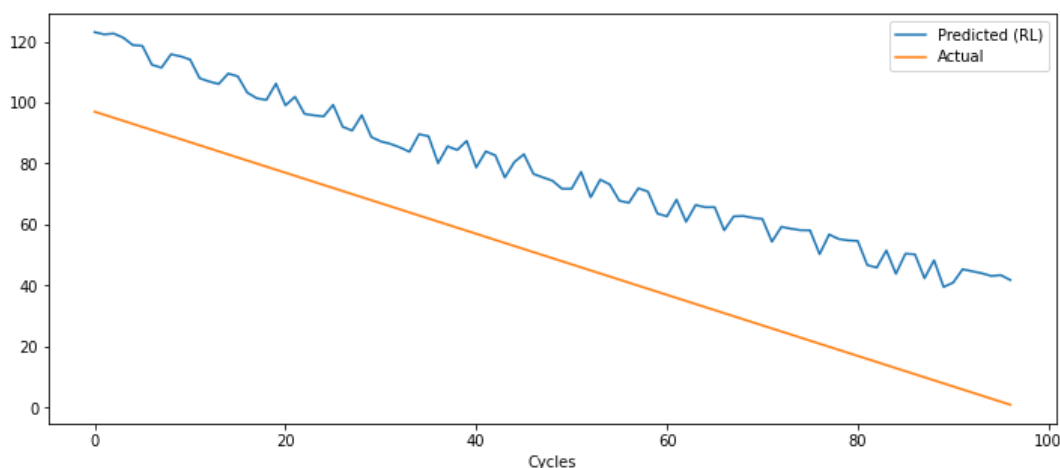


Figura 5.1 RUL previsto vs RUL real - modelo *baseline*

Na Figura 5.1, que representa a comparação entre o RUL previsto com o RUL real, observa-se que existe uma notável distância entre as linhas, evidenciando que o modelo não conseguiu reproduzir, de forma adequada, o comportamento real de degradação da bateria B0018. Contudo, os resultados justificam-se dado que cada uma das baterias do conjunto de dados de treino, atingem o EoL em ciclos diferentes.

5.2.2 RF

O algoritmo RF obteve, em comparação com o modelo *baseline*, um desempenho substancialmente superior, com um RMSE de 9,71, um MAE de 7,65 e um MAPE de 24,69%.

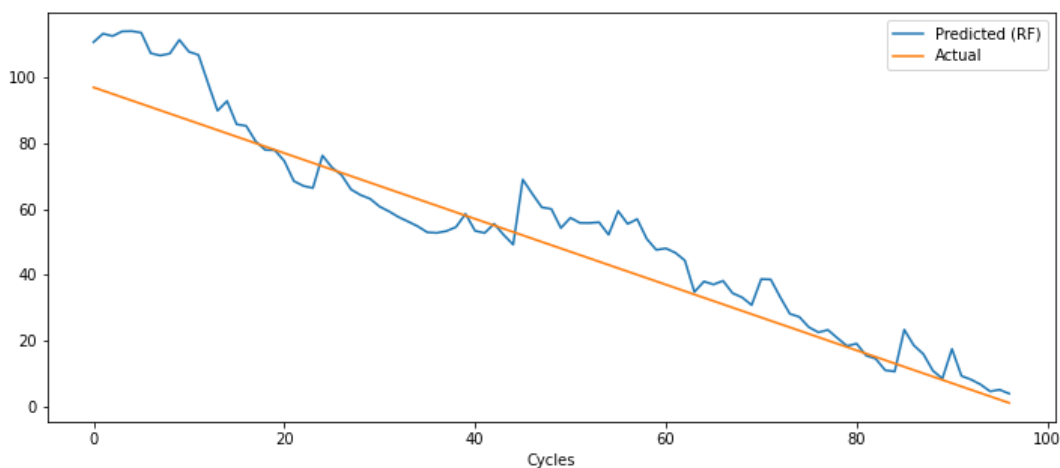


Figura 5.2 RUL previsto vs RUL real - modelo RF

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Através da Figura 5.2, verifica-se uma notável aproximação à linha dos valores do RUL real, o que indica uma boa correspondência entre os valores previstos e os valores reais. Este comportamento valida a capacidade que o algoritmo RF tem em lidar com a complexidade dos dados, explorando as relações não lineares entre as variáveis e diminuindo, de forma significativa, o erro de previsão.

5.2.3 XGBoost

Em comparação com o modelo *baseline* e o RF, o XGBoost obteve resultados intermédios, registando um RMSE de 12,13, um MAE de 10,09 e um MAPE de 25,59%.

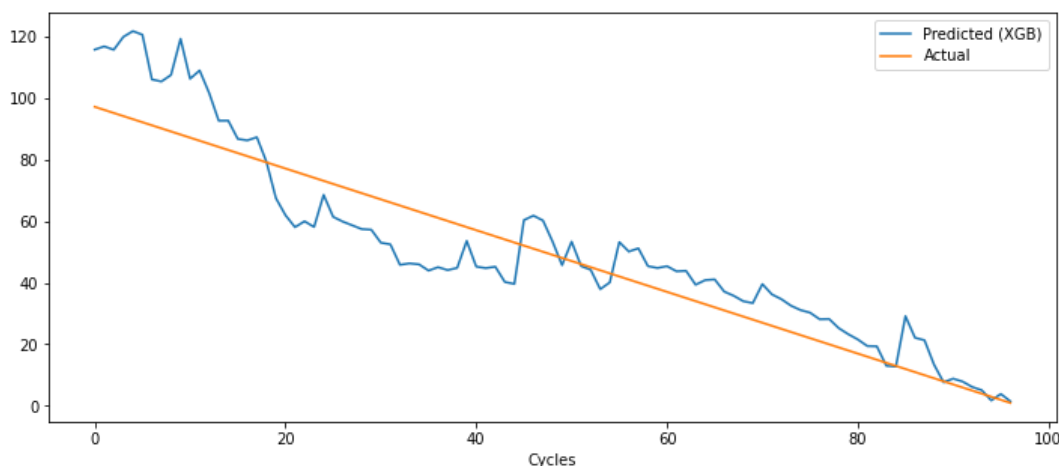


Figura 5.3 RUL previsto vs RUL real - modelo XGBoost

Na Figura 5.3 observa-se que apesar do XGBoost conseguir capturar a tendência global decrescente do RUL real, as previsões feitas apresentam oscilações em torno da linha do RUL real. Contudo, o XGBoost apresenta boa capacidade de previsão, ainda que ligeiramente inferior à do RF.

5.2.4 Comparação dos resultados das métricas

A Tabela 5.1 representa de forma comparativa e resumida os resultados das métricas de avaliação de desempenho do modelo e dos algoritmos.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Tabela 5.1 Comparação das métricas - Iteração 1

Métricas	RL	RF	XGBoost	RL vs. RF	XGBoost vs. RF
RMSE	29,25	9,71	12,13	Melhora	Piora
MAE	28,64	7,65	10,09	Melhora	Piora
MAPE	189,13%	24,69%	25,59%	Melhora	Piora

De modo geral, constatou-se na iteração 1 que o RF apresentou o melhor desempenho global nas três métricas consideradas. O seu RMSE foi significativamente inferior ao dos restantes, o MAE o mais baixo e o MAPE o mais reduzido, refletindo assim, maior precisão e consistência nas previsões do RUL da bateria B0018. Em comparação, o algoritmo XGBoost, mostrou maior variabilidade nos erros e alguma sensibilidade a variações do conjunto de treino. Relativamente ao modelo *baseline*, este apresentou o pior desempenho, o que valida a sua incapacidade de representar processos de degradação de natureza complexa e não linear.

Contudo, pode-se concluir que o algoritmo RF, para a primeira iteração, destacou-se pela sua eficácia na previsão do RUL, apresentando o melhor equilíbrio em termos de capacidade de generalização, precisão e robustez.

5.3 Iteração 2

Na segunda iteração, procedeu-se à otimização dos hiperparâmetros pré-definidos dos algoritmos RF e XGBoost, utilizando as ferramentas *GridSearchCV* e *Optuna*. Posteriormente, os resultados foram comparados e os da ferramenta que apresentou o melhor desempenho foram selecionados. Finalmente, os modelos foram treinados com os "melhores hiperparâmetros" escolhidos, e as suas métricas de desempenho foram comparadas.

5.3.1 Otimização dos hiperparâmetros

GridSearchCV

O *GridSearchCV*, tal como mencionado na secção 4.2, faz uma pesquisa exaustiva aos valores de hiperparâmetros indicados e identifica os valores que melhoram o desempenho dos modelos de previsão.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Os intervalos de valores para cada um dos hiperparâmetros são baseados na Tabela 4.1 para o algoritmo RF e Tabela 4.2 para o algoritmo XGBoost.

Para o modelo RF, foi criada a seguinte grelha de hiperparâmetros:

```
param_grid_rf = {
    'n_estimators': [100, 200, 300, 400, 500],
    'max_depth': [None, 10, 20, 30],
    'max_features': ['sqrt', 'log2'],
    'min_samples_split': [2, 5, 10]
}
```

Figura 5.4 Grelha de hiperparâmetros - RF

onde:

- *n_estimators*: corresponde ao número de árvores na floresta;
- *Max_depth*: corresponde à profundidade máxima da árvore;
- *Max_features*: corresponde ao número de variáveis a considerar por divisão;
- *Min_samples_split*: corresponde ao número mínimo de amostras por nó.

Para o modelo XGBoost, foi criada a seguinte grelha de hiperparâmetros:

```
param_grid_xgb = {
    'n_estimators': [100, 300, 500, 700],
    'learning_rate': [0, 0.01, 0.1, 0.2, 0.3],
    'max_depth': [None, 1, 2, 3, 4, 5, 6],
    'subsample': [0.6, 0.8, 1],
    'booster': ['gbtree']
}
```

Figura 5.5 Grelha de hiperparâmetros - XGBoost

onde:

- *n_estimators*: corresponde ao número de árvores a construir;
- *Learning_rate*: corresponde à redução do tamanho do passo de modo a evitar o sobre ajuste;

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- *Max_depth*: corresponde à profundidade máxima da árvore;
- *Subsample*: corresponde à fração de observações de amostras aleatórias para cada árvore;
- *booster*: corresponde ao tipo de modelo que será usado em cada iteração do processo de *boosting*.

Optuna

O *Optuna* é uma biblioteca para a otimização de hiperparâmetros de modelos de ML. Esta ferramenta, como já referido no Seccção 4.2, é eficiente e permite fazer a otimização dos hiperparâmetros sem a necessidade de recorrer a vários testes de combinações de hiperparâmetros, que por consequência, consomem demasiado tempo.

A Figura 5.6 contém os intervalos para cada um dos hiperparâmetros selecionados para a função objetivo do *Optuna* para o ajuste dos hiperparâmetros do modelo RF:

```
params_rf = {
    'n_estimators': trial.suggest_int('n_estimators', 100, 1000),
    'max_depth': trial.suggest_int('max_depth', 1, 10),
    'max_features': trial.suggest_categorical('max_features', ['sqrt', 'log2']),
    'min_samples_split': trial.suggest_int('min_samples_split', 2, 10)
}
```

Figura 5.6 Intervalo para cada hiperparâmetro - RF

A Figura 5.7 contém os intervalos para cada um dos hiperparâmetros selecionados para a função objetivo do *Optuna* para o ajuste dos hiperparâmetros do modelo XGBoost:

```
params_xgb = {
    'objective': 'reg:squarederror',
    'n_estimators': trial.suggest_int('n_estimators', 100, 1000),
    'learning_rate': trial.suggest_float('learning_rate', 0, 1),
    'max_depth': trial.suggest_int('max_depth', 1, 10),
    'subsample': trial.suggest_float('subsample', 0, 1.0),
    'booster': 'gbtree'
}
```

Figura 5.7 Intervalo para cada hiperparâmetro - XGBoost

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Comparação das ferramentas

Para a seleção da ferramenta de otimização de hiperparâmetros, foram comparados os resultados obtidos da métrica RMSE para a validação cruzada de apenas o conjunto de dados de treino. Um fator adicional considerado para esta comparação foi o tempo de execução, em segundos, que cada ferramenta levou do início ao fim para a otimização dos hiperparâmetros de cada um dos algoritmos. Para tal, recorreu-se à função *time* que permite registar o tempo de processamento por célula.

A Tabela 5.2 representa a comparação das ferramentas para a otimização dos parâmetros para o modelo RF.

Tabela 5.2 *GridSearchCV* vs *Optuna* - RF

Ferramenta	RMSE	Tempo (s)
<i>GridSearchCV</i>	9,35	103,04
<i>Optuna</i>	9,29	113,22

A Tabela 5.3 representa a comparação das ferramentas para a otimização dos hiperparâmetros para o modelo XGBoost.

Tabela 5.3 *GridSearchCV* vs *Optuna* - XGBoost

Ferramenta	RMSE	Tempo (s)
<i>GridSearchCV</i>	7,54	105,26
<i>Optuna</i>	7,44	26,99

A partir das Tabela 5.2 e Tabela 5.3, é possível constatar que o *Optuna* se destacou na otimização dos hiperparâmetros de ambos os algoritmos. Esta ferramenta obteve valores de RMSE e tempos de execução mais reduzidos em comparação com o *GridSearchCV*, à exceção da otimização dos hiperparâmetros do RF.

5.3.2 Treinar o algoritmo RF Otimizado

A Tabela 5.1 representa a comparação entre os resultados das métricas de avaliação de desempenho do RF, da iteração 1 e o RF Otimizado, com recurso à ferramenta *Optuna*.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Tabela 5.4 RF vs RF Otimizado

Métricas	RF	RF otimizado	RF Otimizado vs. RF
RMSE	9,71	9,64	Melhora
MAE	7,65	7,61	Melhora ligeiramente
MAPE	24,69%	26,04%	Piora

Com a otimização dos hiperparâmetros, o RF apresentou melhorias nas métricas RMSE e MAE. O RMSE reduziu de 9,71 para 9,64, e o MAE de 7,65 para 7,61. No entanto, o MAPE teve um aumento, registando uma subida de 24,69% para 26,04%.

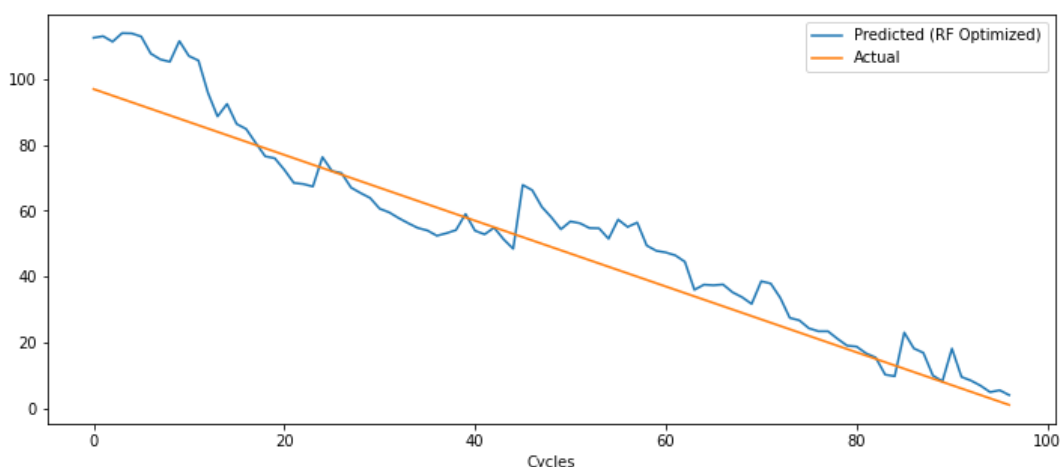


Figura 5.8 RUL previsto vs RUL real - RF Otimizado

Na Figura 5.8 observa-se que, apesar das oscilações, a linha dos RUL previsto tende a aproximar-se à linha do RUL real, o que indica uma redução das divergências entre os valores previstos e reais. Posto isto, e em termos gerais, a otimização permitiu ao algoritmo captar as variações no comportamento de degradação da bateria de forma mais eficiente.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

5.3.3 Treinar o algoritmo XGBoost Otimizado

A Tabela 5.1 representa a comparação entre os resultados das métricas de avaliação de desempenho do XGBoost e XGBoost Otimizado.

Tabela 5.5 XGBoost vs XGBoost Otimizado

Métricas	XGBoost	XGBoost otimizado	XGBoost Otimizado vs. XGBoost
RMSE	12,13	14,28	Piora
MAE	10,09	11,98	Piora
MAPE	25,59%	40,81%	Piora

Após a otimização dos hiperparâmetros do XGBoost constatou-se que o algoritmo não apresentou melhoria dos resultados. O RMSE aumentou de 12,13 para 14,28, e o MAE de 10,09 para 11,98, e o MAPE de 25,59% para 40,81%. Este aumento resultou na redução da capacidade de generalização do algoritmo.

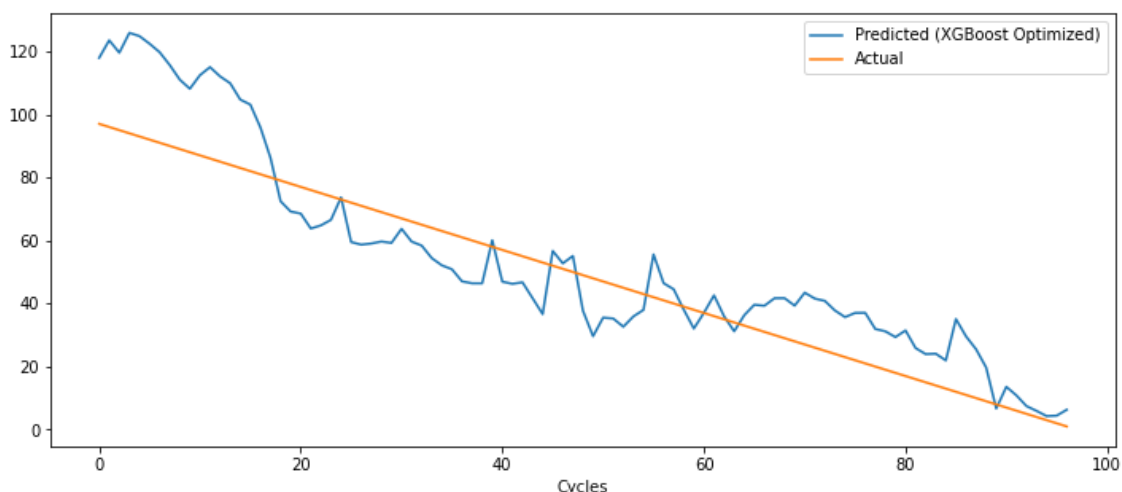


Figura 5.9 RUL previsto vs RUL real - XGBoost Otimizado

Verifica-se, na Figura 5.9, que fica evidente que o XGBoost Otimizado não tem uma boa precisão preditiva, existindo uma dispersão entre as linhas do RUL previsto e do RUL real, quando comparado com o algoritmo não otimizado.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

5.3.4 Comparação entre os algoritmos antes e depois da otimização

A otimização dos hiperparâmetros com recurso à utilização da ferramenta *Optuna*, contribuiu para uma ligeira melhoria nas previsões do RUL para o RF. Por outro lado, a otimização do XGBoost mostrou que o algoritmo tem algumas dificuldades em prever o RUL real, algo notável nos primeiros valores do RUL, ficando evidente que houve perda da capacidade de generalização.

Contudo, o RF continua a ser o algoritmo que, em termos globais, tem o melhor desempenho, mantendo os resultados das métricas de desempenho mais baixos em comparação com o XGBoost.

5.4 Iteração 3

Com o intuito de melhorar a capacidade de generalização do modelo *baseline* e dos algoritmos RF e XGBoost, e reduzir as diferenças que existem entre os perfis de RUL das baterias, foi realizada uma terceira iteração. Nesta iteração, procedeu-se à normalização da variável RUL através da função *minmaxscaler*, que foi aplicada apenas ao conjunto de treino. Procedeu-se desta forma porque assume-se que a bateria B0018 é uma bateria a qual não se conhece o perfil de degradação.

Seguindo o mesmo princípio da iteração 1, foram novamente treinados o modelo *baseline* e os algoritmos e posteriormente comparados os seus resultados com os algoritmos otimizados na iteração 2 e com o modelo *baseline*.

5.4.1 Modelo *baseline*

Para o modelo *baseline*, a normalização da variável RUL não teve qualquer impacto no desempenho do modelo, mantendo-se as métricas idênticas às observadas na primeira iteração, ou seja, $RMSE = 29,25$, $MAE = 28,64$ e $MAPE = 189,13\%$.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

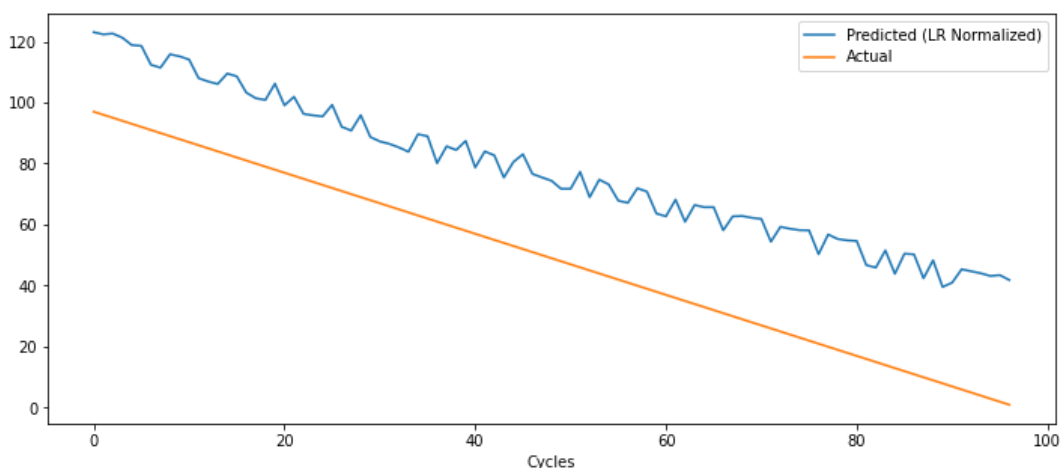


Figura 5.10 RUL previsto vs RUL real normalizado - modelo *baseline*

A Figura 5.10 ilustra o mesmo padrão de dispersão das linhas, indicando mais uma vez que o modelo não conseguiu representar, de forma adequada, as relações não lineares do processo de degradação. Posto isto, a normalização do RUL não trouxe benefícios adicionais ao modelo *baseline*.

5.4.2 RF

Após a normalização da variável RUL, o RF obteve os seguintes resultados: RMSE=9,71, MAE=7,84 e MAPE=26,82%. Em comparação com o algoritmo otimizado da segunda iteração, observa-se um ligeiro aumento nos erros quadrados e absolutos, o que sugere uma ligeira perda de precisão da previsão.

Torna-se evidente na Figura 5.11, que a dispersão entre as linhas é maior, validando assim a ligeira perda de precisão. Contudo, o RF manteve uma boa coerência e estabilidade nos resultados.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

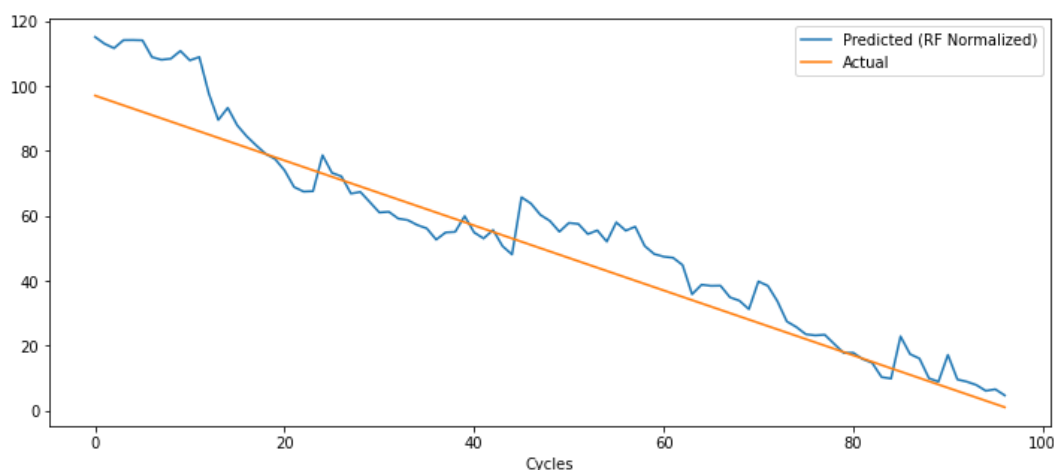


Figura 5.11 RUL previsto vs RUL real normalizado - algoritmo RF

5.4.3 XGBoost

O XGBoost apresentou um RMSE de 11,43, um MAE de 8,98 e um MAPE de 25,44%, com a normalização do RUL. Em comparação com o algoritmo otimizado, é evidente houve uma redução nos resultados das métricas, indicando uma melhoria na capacidade de generalização.

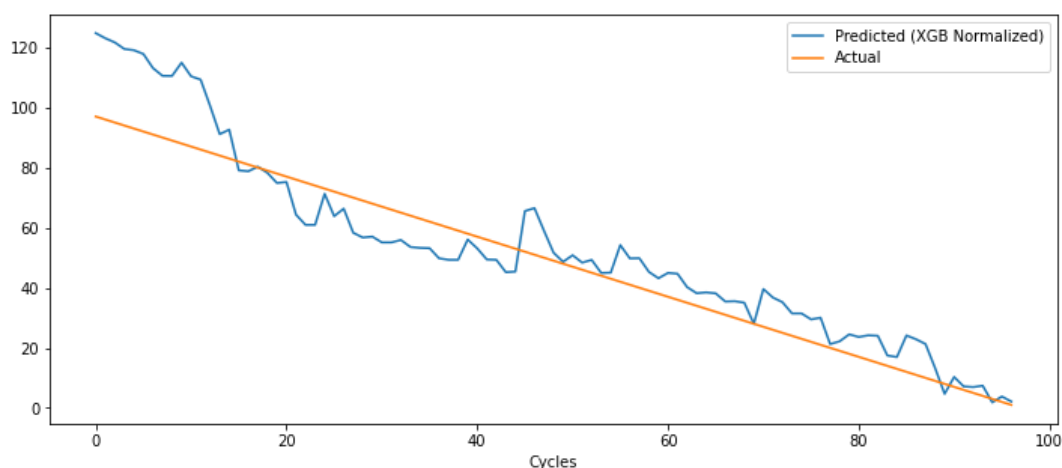


Figura 5.12 RUL previsto vs RUL real normalizado - algoritmo XGBoost

A partir da Figura 5.12 observa-se que o XGBoost consegue captar a tendência global da decrescente do RUL real, tendo a linha do RUL previsto mais próxima à linha do RUL real. Contudo, o RUL normalizado, trouxe melhorias significativas para o algoritmo XGBoost.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

5.4.4 Comparação dos resultados das métricas dos modelos

A Tabela 5.6 contém os resultados comparativos das métricas de avaliação de desempenho antes e após a normalização da variável RUL.

Tabela 5.6 Comparação das métricas dos modelos - Iteração 3

Métricas	RL	RL - RUL normalizado	RF otimizado	RF - RUL normalizado	XGBoost otimizado	XGBoost RUL normalizado
RMSE	29,25	29,25	9,64	9,71	14,28	11,43
MAE	28,64	28,64	7,61	7,84	11,98	8,98
MAPE	189,13%	189,13%	26,04%	26,82%	40,81%	25,44%

Com a normalização do RUL, os algoritmos RF e XGBoost, à semelhança dos resultados obtidos na iteração 1, superaram o modelo *baseline* em todas as métricas. Isto confirma que as relações que existem entre as variáveis de independentes e o RUL, não são lineares.

Entre os dois algoritmos, RF e XGBoost, a normalização do RUL trouxe benefícios notáveis para o XGBoost, melhorando a capacidade de generalização. O resultado do MAE, na terceira iteração foi o que registou o melhor MAE em comparação com os resultados da mesma métrica nas duas primeiras iterações.

Em suma, os resultados obtidos comprovam que o RF e o XGBoost são adequados para a previsão do RUL, sendo que quando se trata de otimização dos hiperparâmetros, o RF destacou-se como o melhor modelo para a previsão do RUL. Contudo e de um modo geral, o XGBoost com o RUL normalizado destaca-se como o algoritmo com o equilíbrio entre precisão e robustez.

CONCLUSÃO

1. Pequeno sumário do que foi feito

A presente dissertação passou pelos estudos da previsão do RUL das baterias de iões de lítio a partir de conjuntos de dados experimentais, nomeadamente *BatteryAgingARC-FY08Q4*, tendo como foco cenários de generalização para baterias as quais são desconhecidos os padrões de degradação ou que se encontram no seu estado novo. A previsão do RUL das baterias é um aspeto crítico para a implementação da estratégia de manutenção preditiva dos BEV.

Após a preparação dos dados e a extração de variáveis relevantes para a previsão do RUL, foram treinados e avaliados três modelos de ML: RL, RF e XGBoost. O estudo decorreu em três iterações, criadas para perceber se existe alguma melhoria nos resultados da previsão em cada uma das iterações. Na primeira iteração, estabeleceu-se a linha de base do desempenho, caracterizando a capacidade preditiva de cada modelo face ao problema em análise. Na segunda, otimizaram-se o RF e o XGBoost, comparando duas estratégias, *GridSearchCV* e *Optuna*, e selecionando, para cada algoritmo, a estratégia que melhor otimizou os hiperparâmetros dos algoritmos. Na terceira e última iteração, aplicou-se a normalização do RUL exclusivamente ao conjunto de treino, e procedeu-se à comparação sistemática das métricas obtidas nas três iterações, com o objetivo de verificar eventuais melhorias de desempenho.

2. Síntese dos principais resultados

Os resultados obtidos em cada uma das iterações, revelam diferenças significativas no desempenho dos três modelos. Na primeira iteração, RF foi o modelo com os melhores resultados das métricas de avaliação de desempenho, (RMSE de 9,71, MAE de 7,65 e MAPE de 24,69%), seguido do XGBoost que apresentou pequenas diferenças nos resultados em comparação. O modelo *baseline*, teve o pior desempenho, tendo resultados significativamente superiores o que indica que o modelo não é adequado para a complexidade da previsão do RUL de uma bateria desconhecida.

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Na iteração 2, constatou-se que a otimização trouxe resultados mistos. Por um lado, O RF Otimizado, manteve um desempenho robusto de previsão. Por outro lado, o XGBoost Otimizado registou uma notável degradação nos resultados das métricas de desempenho (RMSE de 14,28, MAE de 11,98 e MAPE de 40,81%), o que indica que o algoritmo teve uma perda da capacidade de generalização.

Contudo, a normalização do RUL na iteração 3, demonstrou ser um fator crucial. Embora que não se tenha verificado nenhuma alteração ao desempenho do modelo *baseline*, e tenha sido tido um impacto estatisticamente irrelevante no desempenho do RF, a normalização do RUL teve impactos positivos para o desempenho do XGBoost. Após a normalização, o XGBoost alcançou uma melhoria no desempenho (RMSE de 11,43, MAE de 8,98 e MAPE de 25,44%), superando inclusivamente o RF em termos de MAE e equiparando-se em RMSE, o que o posiciona como o melhor algoritmo para a previsão do RUL de baterias desconhecidas quando o RUL é normalizado.

Limitações

A principal limitação é que estudo teve foi não considerar um conjunto de dados para a validação. Ou seja, de modo a otimizar melhor os modelos, o ideal seria usar duas baterias para o conjunto de treino e uma para a validação. A falta deste passo fundamental, deixa evidente em particular para o XGBoost que registou uma degradação nos resultados das métricas de avaliação.

Ainda em relação ao processo de otimização dos hiperparâmetros, e uma vez que os hiperparâmetros otimizados selecionados foram os do *Optuna*, importa referir que durante a otimização não foi ativado o mecanismo *early stopping* (interrupção antecipada do treino). A decisão da não ativação visou a garantir que todas as combinações de hiperparâmetros fossem treinadas até o limite definido, de modo a preservar a comparabilidade com a iteração 1.

Trabalhos futuros

Com base nos resultados obtidos com o estudo e nas limitações identificadas, propõe-se a otimização dos hiperparâmetros através da utilização dos dados de uma bateria para a

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

validação, ficando o conjunto de treino com duas baterias. Este passo contribuirá para a melhoria do desempenho do modelo e para sua capacidade de generalização em particular quando se trata de previsão de baterias cujo padrão de degradação é desconhecido. Adicionalmente, propõem-se a utilização de modelos de previsão híbridos, ou seja, juntar dois ou mais modelos que poderá oferecer melhorias significativas em relação à previsão do RUL.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abiodun, A. A., Toshio, A., & Kamil, S. K. (2024). Prediction Model for Electricity Energy Consumption and Pricing Using Xgboost With L1 Regularization. *Kasu Journal of Computer Science*, 1(3), 640–652. <https://doi.org/10.47514/kjcs/2024.1.3.0017>
- Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. In *Applied Sciences (Switzerland)* (Vol. 12, Issue 16). MDPI. <https://doi.org/10.3390/app12168081>
- Ahmadian, A., Sedghi, M., Elkamel, A., Fowler, M., & Aliakbar Golkar, M. (2018). Plug-in electric vehicle batteries degradation modeling for smart grid studies: Review, assessment and conceptual framework. In *Renewable and Sustainable Energy Reviews* (Vol. 81, pp. 2609–2624). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2017.06.067>
- Amazon Web Services, Inc. (2025). What is hyperparameter tuning? <https://aws.amazon.com/what-is/hyperparameter-tuning/>. (Acedido a 2025-10-16)
- Ansari, S., Ayob, A., Lipu, M. S. H., Hussain, A., & Saad, M. H. M. (2021). Data-driven remaining useful life prediction for lithium-ion batteries using multi-charging profile framework: A recurrent neural network approach. *Sustainability (Switzerland)*, 13(23). <https://doi.org/10.3390/su132313333>
- Arena, F., Collotta, M., Luca, L., Ruggieri, M., & Termine, F. G. (2022). Predictive Maintenance in the Automotive Sector: A Literature Review. *Mathematical and Computational Applications*, 27(1), 2. <https://doi.org/10.3390/mca27010002>
- Atalay, S., Sheikh, M., Mariani, A., Merla, Y., Bower, E., & Widanage, W. D. (2020). Theory of battery ageing in a lithium-ion battery: Capacity fade, nonlinear ageing and lifetime prediction. *Journal of Power Sources*, 478, 229026. <https://doi.org/10.1016/j.jpowsour.2020.229026>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- Chang, Y. H., Hsieh, Y. C., Chai, Y. H., & Lin, H. W. (2023). Remaining-Useful-Life Prediction for Li-Ion Batteries. *Energies*, 16(7). <https://doi.org/10.3390/en16073096>
- Chen, J. C., Chen, T. L., Liu, W. J., Cheng, C. C., & Li, M. G. (2021). Combining empirical mode decomposition and deep recurrent neural networks for predictive maintenance of lithium-ion battery. *Advanced Engineering Informatics*, 50. <https://doi.org/10.1016/j.aei.2021.101405>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, Z., Chen, L., Shen, W., & Xu, K. (2022). Remaining Useful Life Prediction of Lithium-Ion Battery via a Sequence Decomposition and Deep Learning Integrated Approach. *IEEE Transactions on Vehicular Technology*, 71(2), 1466–1479. <https://doi.org/10.1109/TVT.2021.3134312>
- Çinar, Z. M., Nuhu, A. A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability (Switzerland)*, 12(19). <https://doi.org/10.3390/su12198211>
- Deng, Y., & Lumley, T. (2023). *Multiple Imputation Through XGBoost*. <http://arxiv.org/abs/2106.01574>
- dos Reis, G., Strange, C., Yadav, M., & Li, S. (2021). Lithium-ion battery data and where to find it. *Energy and AI*, 5. <https://doi.org/10.1016/j.egyai.2021.100081>
- Elmahallawy, M., Elfouly, T., Alouani, A., & Massoud, A. M. (2022). A Comprehensive Review of Lithium-Ion Batteries Modeling, and State of Health and Remaining Useful Lifetime Prediction. *IEEE Access*, 10, 119040–119070. <https://doi.org/10.1109/ACCESS.2022.3221137>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- Ezzouhri, A., Charouh, Z., Ghogho, M., & Guennoun, Z. (2023). A Data-Driven-Based Framework for Battery Remaining Useful Life Prediction. *IEEE Access*, 11, 76142–76155. <https://doi.org/10.1109/ACCESS.2023.3286307>
- Fanoro, M., Božanić, M., & Sinha, S. (2022). A Review of the Impact of Battery Degradation on Energy Management Systems with a Special Emphasis on Electric Vehicles. In *Energies* (Vol. 15, Issue 16). MDPI. <https://doi.org/10.3390/en15165889>
- Feng, J., Cai, F., Li, H., Huang, K., & Yin, H. (2023). A data-driven prediction model for the remaining useful life prediction of lithium-ion batteries. *Process Safety and Environmental Protection*, 180, 601–615. <https://doi.org/10.1016/j.psep.2023.10.042>
- Hasib, S. A., Islam, S., Chakraborty, R. K., Ryan, M. J., Saha, D. K., Ahamed, M. H., Moyeen, S. I., Das, S. K., Ali, M. F., Islam, M. R., Tasneem, Z., & Badal, F. R. (2021). A Comprehensive Review of Available Battery Datasets, RUL Prediction Approaches, and Advanced Battery Management. *IEEE Access*, 9, 86166–86193. <https://doi.org/10.1109/ACCESS.2021.3089032>
- Hurtado, J., Salvati, D., Semola, R., Bosio, M., & Lomonaco, V. (2023). Continual learning for predictive maintenance: Overview and challenges. In *Intelligent Systems with Applications* (Vol. 19). Elsevier B.V. <https://doi.org/10.1016/j.iswa.2023.200251>
- Jafari, S., & Byun, Y. C. (2022). XGBoost-Based Remaining Useful Life Estimation Model with Extended Kalman Particle Filter for Lithium-Ion Batteries. *Sensors*, 22(23). <https://doi.org/10.3390/s22239522>
- Jafari, S., Byun, Y. C., & Ko, S. (2023). A Novel Approach for Predicting Remaining Useful Life and Capacity Fade in Lithium-Ion Batteries Using Hybrid Machine Learning. *IEEE Access*, 11, 131950–131963. <https://doi.org/10.1109/ACCESS.2023.3329508>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- Jafari, S., Yang, J. H., & Byun, Y. C. (2024). Optimized XGBoost modeling for accurate battery capacity degradation prediction. *Results in Engineering*, 24. <https://doi.org/10.1016/j.rineng.2024.102786>
- Jia, Z., Wang, Z., Sun, Z., Liu, P., Zhu, X., & Sun, F. (2023). A Data-Driven Approach for Battery System Safety Risk Evaluation Based on Real-world Electric Vehicle Operating Data. *IEEE Transactions on Transportation Electrification*. <https://doi.org/10.1109/TTE.2023.3324450>
- Kiriakidou, N., Livieris, I. E., & Diou, C. (2024). C-XGBoost: A tree boosting model for causal effect estimation. <http://arxiv.org/abs/2404.00751>
- Li, L., Li, Y., Mao, R., Li, L., Hua, W., & Zhang, J. (2023). Remaining Useful Life Prediction for Lithium-Ion Batteries With a Hybrid Model Based on TCN-GRU-DNN and Dual Attention Mechanism. *IEEE Transactions on Transportation Electrification*, 9(3), 4726–4740. <https://doi.org/10.1109/TTE.2023.3247614>
- Li, R., Sun, H., Wei, X., Ta, W., & Wang, H. (2022). Lithium Battery State-of-Charge Estimation Based on AdaBoost.Rt-RNN. *Energies*, 15(16). <https://doi.org/10.3390/en15166056>
- Li, X., Ju, L., Geng, G., & Jiang, Q. (2023). Data-driven state-of-health estimation for lithium-ion battery based on aging features. *Energy*, 274. <https://doi.org/10.1016/j.energy.2023.127378>
- Lin, M., Yan, C., Wang, W., Dong, G., Meng, J., & Wu, J. (2023). A data-driven approach for estimating state-of-health of lithium-ion batteries considering internal resistance. *Energy*, 277. <https://doi.org/10.1016/j.energy.2023.127675>
- Liu, Y., Sun, G., & Liu, X. (2023). Remaining useful life prediction of lithium-ion batteries based on peak interval features and deep learning. *Journal of Energy Storage*, 73. <https://doi.org/10.1016/j.est.2023.109308>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- Nunes, P., Santos, J., & Rocha, E. (2023). Challenges in predictive maintenance – A review. In *CIRP Journal of Manufacturing Science and Technology* (Vol. 40, pp. 53–67). Elsevier Ltd. <https://doi.org/10.1016/j.cirpj.2022.11.004>
- scikit-learn developers. (s.d.). GridSearchCV — scikit-learn 1.7.2 documentation. In scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. (Acedido a 2025-10-16)
- scikit-learn developers. (s.d.). RandomForestRegressor. In scikit-learn: Machine Learning in Python (Version 1.5). scikit-learn documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>. (Acedido a 2025-10-16)
- scikit-learn developers. (s.d.). SelectKBest — scikit-learn 1.7.2 documentation. In scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html. (Acedido a 2025-10-17)
- Sharma, P., & Bora, B. J. (2023). A Review of Modern Machine Learning Techniques in the Prediction of Remaining Useful Life of Lithium-Ion Batteries. In *Batteries* (Vol. 9, Issue 1). MDPI. <https://doi.org/10.3390/batteries9010013>
- Theissler, A., Pérez-Velázquez, J., Kettelgerdes, M., & Elger, G. (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. *Reliability Engineering and System Safety*, 215. <https://doi.org/10.1016/j.ress.2021.107864>
- Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. In *Signal Processing* (Vol. 167). Elsevier B.V. <https://doi.org/10.1016/j.sigpro.2019.107299>
- Wang, G., Lyu, Z., & Li, X. (2023). An Optimized Random Forest Regression Model for Li-Ion Battery Prognostics and Health Management. *Batteries*, 9(6). <https://doi.org/10.3390/batteries9060332>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

- Wang, Z., Liu, Y., Wang, F., Wang, H., & Su, M. (2023). Capacity and remaining useful life prediction for lithium-ion batteries based on sequence decomposition and a deep-learning network. *Journal of Energy Storage*, 72. <https://doi.org/10.1016/j.est.2023.108085>
- Warni, E., Saputri, D. W., Prayogi Alimuddin, A. A., Muhammad Rizal, H., Arif, Y. M., & Rahman, I. T. (2024). Boosting Methods Comparison : XGBoost, Adaboost and Gradient Boosting for Business Partner Performance Prediction. *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2024*, 376–380. <https://doi.org/10.1109/ICITISEE63424.2024.10730221>
- Wu, J., Liu, Z., Zhang, Y., Lei, D., Zhang, B., & Cao, W. (2023). Data-driven state of health estimation for lithium-ion battery based on voltage variation curves. *Journal of Energy Storage*, 73. <https://doi.org/10.1016/j.est.2023.109191>
- XGBoost Developers. (s.d.). XGBoost parameters. XGBoost Documentation. <https://xgboost.readthedocs.io/en/stable/parameter.html>. (Acedido a 2025-10-17)
- Xiao, Z., Jiang, B., Zhu, J., Wei, X., & Dai, H. (2024). State of Health Estimation for Lithium-Ion Batteries Using an Explainable XGBoost Model with Parameter Optimization. *Batteries*, 10(11). <https://doi.org/10.3390/batteries10110394>
- Xu, H., Yan, S., Qin, X., An, W., & Wang, J. (2025). Interpretable Intelligent Fault Diagnosis for Heat Exchangers Based on SHAP and XGBoost. *Processes*, 13(1). <https://doi.org/10.3390/pr13010219>
- Zhao, J., Zhu, Y., Zhang, B., Liu, M., Wang, J., Liu, C., & Hao, X. (2023). Review of State Estimation and Remaining Useful Life Prediction Methods for Lithium-Ion Batteries. In *Sustainability (Switzerland)* (Vol. 15, Issue 6). MDPI. <https://doi.org/10.3390/su15065014>
- Zhu, C., He, Z., Bao, Z., Sun, C., & Gao, M. (2023). Prognosis of Lithium-Ion Batteries' Remaining Useful Life Based on a Sequence-to-Sequence Model

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

with Variational Mode Decomposition. *Energies*, 16(2).

<https://doi.org/10.3390/en16020803>

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

APÊNDICES

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

APÊNDICE 1. Estatísticas descritivas

Tabela 0.1 Estatísticas descritivas da bateria B0007

Variáveis	Total	Média	Desvio Padrão	Mínimo	Máximo
Ciclo	591458	-	-	1	338
Carga	541173	-	-	1	170
Descarga	50285	-	-	1	168
Temperatura ambiente	591458	24	0	24	24
Timestamp	591458	-	-	02/04/2008 13:08:18	28/05/2008 11:09:42
Voltagem medida	591458	4,1	0,22	0	8,33
Corrente medida	591458	0,39	0,93	-2,26	1,51
Temperatura medida	591458	26,12	2,72	22,97	42,33
Corrente na fonte de carregamento	591458	0,68	0,75	-2,27	2
Voltagem na fonte de carregamento	591458	3,87	1,43	0	5
Tempo em segundos	591458	4763,86	3147,86	0	10807,33
Capacidade	591453	1,61	0,14	1,41	1,89

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

Tabela 0.2 Estatísticas descritivas da bateria B0018

Variáveis	Total	Média	Desvio Padrão	Mínimo	Máximo
Ciclo	314676	-	-	1	266
Carga	279810	-	-	1	134
Descarga	34866	-	-	1	132
Temperatura ambiente	314676	24	0	24	24
Timestamp	314676	-	-	07/07/2008 12:26:46	20/08/2008 08:37:20
Voltagem medida	314674	4,08	0,24	2,28	4,3
Corrente medida	314674	0,31	0,98	-4,51	1,54
Temperatura medida	314674	25,91	2,99	21,81	38,88
Corrente na fonte de carregamento	314676	0,71	0,73	-4,51	2
Voltagem na fonte de carregamento	314676	4,02	1,13	-0,01	5
Tempo em segundos	314676	4722,18	3111,16	0	10815,38
Capacidade	312497	1,6	0,16	1,34	1,86

Previsão da Vida Útil Restante para a Manutenção Preditiva das Baterias de Veículos Elétricos: Um Contributo com Conjuntos de Dados Benchmark

APÊNDICE 2. Pipeline da modelação

