Modelos de Machine Learning para a Deteção de Anomalias em Consumos Energéticos

Mestrado em Ciência de Dados

Beatriz Colaço dos Santos

Leiria, abril de 2026

Modelos de Machine Learning para a Deteção de Anomalias em Consumos Energéticos

Mestrado em Ciência de Dados

Beatriz Colaço dos Santos

Projeto de Mestrado realizado sob a orientação do Professor Doutor Rolando Miragaia, do Professor Doutor Carlos Grilo e do Professor Doutor Fernando Sebastião, docentes do Instituto Politécnico de Leiria.

Leiria, abril de 2026

Originalidade e Direitos de Autor

O presente relatório de projeto é original, elaborado unicamente para este fim, tendo sido devidamente citados todos os autores cujos estudos e publicações contribuíram para o elaborar.

Agradecimentos

Antes de mais, gostaria de expressar o meu sincero agradecimento aos meus orientadores, Professor Carlos Grilo, Professor Rolando Miragaia e Professor Fernando Sebastião. Apesar das diversas mudanças pessoais e profissionais que ocorreram ao longo deste percurso, estiveram sempre disponíveis para prestar apoio, orientação e incentivo, contribuindo de forma decisiva para a concretização deste trabalho.

À minha família, mãe, pai e irmã, que estiveram sempre presentes, mesmo nos dias menos bons, oferecendo sempre uma palavra de apoio e incentivo. Ao Nuno, por nunca me deixar desistir e por acreditar em mim ao longo de todo este caminho.

E, por fim, à Susana, à Diana e ao Alexandre, pelas amizades que nunca pensei criar durante o mestrado e que sei que levarei comigo para sempre.

Resumo

O consumo de energia em edifícios institucionais é influenciado por múltiplos fatores e pode ser difícil de monitorizar de forma eficiente sem ferramentas adequadas de apoio à decisão. A previsão do consumo energético, quando realizada com elevada precisão, pode ser utilizada como referência de comportamento esperado, permitindo identificar desvios significativos que possam corresponder a anomalias operacionais. O presente trabalho, desenvolvido em continuidade com o trabalho de investigação de J. Sá, tem como objetivo o desenvolvimento de modelos de estimação do consumo diário de energia com um erro reduzido, de modo a poderem ser utilizados como suporte à deteção de anomalias no consumo energético. A metodologia seguida consiste na utilização de um modelo de previsão horária do consumo e na agregação dos valores previstos ao longo do dia, permitindo obter uma estimativa diária do consumo. Os resultados mostram que esta metodologia simples permite obter valores de erro extremamente baixos, deixando em aberto a possibilidade da sua utilização como suporte à deteção de anomalias no consumo. Foram utilizados os registos do Campus do Instituto Politécnico de Leiria, recolhidos entre 2015 e 2023, para avaliar o desempenho da metodologia proposta. Para tal, foram desenvolvidos e comparados seis modelos: MLP, LSTM, GRU, XGBoost, N-BEATS e N-HiTS. Os resultados demonstram que esta metodologia permite obter erros com MAPE médio entre 1,28% e 1,70%, evidenciando o bom desempenho preditivo dos modelos desenvolvidos.

Palavras-chave: consumo energético, estimação do consumo, séries temporais, *machine learning*, deteção de anomalias

Abstract

Energy consumption in institutional buildings is influenced by multiple factors and can be challenging to monitor efficiently without adequate decision-support tools. Energy consumption forecasting, when performed with high accuracy, can serve as a reference for expected behavior, enabling the identification of significant deviations that may correspond to operational anomalies. The present study, developed as a continuation of the research conducted by J. Sá, aims to develop energy consumption estimation models with low error, so that they can be used to support anomaly detection in energy consumption. The methodology consists of using an hourly consumption forecasting model and aggregating the predicted values throughout the day, yielding a daily consumption estimate. The results show that this simple methodology achieves very low error values, indicating its potential use as support for anomaly detection in energy consumption. Hourly records from the Campus of the Polytechnic Institute of Leiria, collected between 2015 and 2023, were used to evaluate the performance of the proposed methodology. Six models were developed and compared: MLP, LSTM, GRU, XGBoost, N-BEATS and N-HiTS. The results demonstrate that this methodology achieves mean MAPE values between 1,28% and 1,70%, indicating the good predictive performance of the developed models.

Keywords: energy consumption, energy consumption estimation, time series, machine learning, anomaly detection

Índice

Lista de Figuras

Lista de tabelas

Lista de siglas e acrónimos

ARIMA	AutoRegressive Integrated Moving Average
BPTT	Backpropagation Through Time
CAE	Convolutional Autoencoder
CRISP-DM	Cross-Industry Standard Process for Data Mining
DL	Deep Learning
DT	Decision Tree
GBR	Gradient Boosting Regressor
GNN	Graph Neural Networks
GRU	Gated Recurrent Unit
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MFDA	Modified Flow Direction Algorithm
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MSE	Mean Squared Error
N-BEATS	Neural Basis Expansion Analysis for Time Series
N-HiTS	Neural Hierarchical Interpolation for Time Series
R^2	Coeficiente de Determinação
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SARIMA	Seasonal AutoRegressive Integrated Moving Average
SGD	Stochastic Gradient Descent
SVR	Support Vector Regression
TCN	Temporal Convolutional Network
XGBoost	eXtreme Gradient Boosting

1. Introdução

O consumo de energia é uma constante na atividade humana, estando presente de forma contínua tanto no contexto pessoal como profissional. O aumento da dependência energética, aliado às preocupações ambientais e aos custos associados, reforça a necessidade de uma utilização mais eficiente e sustentável dos recursos energéticos [1]. Neste contexto, torna-se essencial que organizações e instituições implementem estratégias para monitorizar e controlar o consumo, com o objetivo de reduzir desperdícios, otimizar a utilização da energia e mitigar os impactos ambientais associados [1, 2].

A previsão do consumo energético tem sido amplamente estudada na literatura, sendo aplicada no planeamento e na otimização dos sistemas energéticos [3]. No entanto, a previsão futura, por si só, não constitui um critério suficiente para determinar se o consumo observado num determinado instante corresponde a um comportamento normal ou anómalo [4]. Para uma gestão energética operacional eficaz, torna-se essencial comparar o consumo observado com um consumo esperado de referência, permitindo identificar desvios significativos. Neste contexto, os modelos preditivos podem ser utilizados como base para estimar este consumo esperado, suportando a deteção de anomalias através da análise dos desvios (resíduos) entre valores previstos e observados [5].

Assim, o presente trabalho propõe o desenvolvimento de modelos de estimação do consumo diário de energia, com erro reduzido, com vista à sua utilização como suporte à deteção de anomalias no consumo energético em edifícios institucionais. Para tal, recorreu-se a algoritmos de *machine learning* (ML) aplicados a séries temporais de consumo, sendo a estimativa diária obtida por agregação das previsões horárias geradas pelos modelos desenvolvidos.

1.1. Objetivos

O presente trabalho dá continuidade à investigação desenvolvida por J. Sá [6] no âmbito do Mestrado em Ciência de Dados no Instituto Politécnico de Leiria, na qual foi estudada a previsão do consumo energético em edifícios institucionais. Nesse trabalho, o autor sugeriu a utilização dos modelos de previsão desenvolvidos como referência de consumo esperado, permitindo a comparação a *posteriori* entre valores previstos e observados, com vista à

identificação de desvios significativos no consumo energético, potencialmente associados a comportamentos anómalos. A abordagem assenta no desenvolvimento de um modelo de ML capaz de prever com elevada precisão o consumo energético ao longo de um horizonte temporal de 24 horas. Uma vez alcançado esse nível de desempenho, torna-se possível recorrer ao modelo para detetar anomalias no consumo real, através da comparação entre o consumo previsto (que representa o consumo esperado em condições normais de funcionamento) e o consumo efetivamente observado.

O objetivo desta dissertação é o desenvolvimento de modelos de estimação de consumo energético diário com erro reduzido, para suporte à deteção de comportamentos anómalos, através da análise dos desvios entre valores previstos e observados em edifícios institucionais. O presente trabalho centra-se na obtenção de modelos com elevado desempenho, avaliados em termos de erro de previsão, não sendo abordada a implementação operacional da deteção de anomalias.

De forma a concretizar o objetivo geral, definem-se os seguintes objetivos específicos:

- Desenvolver e ajustar diferentes modelos de estimação do consumo energético diário com base em séries temporais históricas de consumo energético horário;
- Avaliar o desempenho dos modelos de estimação com recurso a métricas de erro adequadas;
- Analisar os desvios (resíduos) entre os valores estimados e os valores efetivamente observados;
- Avaliar a robustez e consistência dos modelos através da realização de múltiplos ciclos de treino independentes.

1.2. Metodologia

A metodologia adotada neste trabalho assenta nos princípios do CRISP-DM (*Cross-Industry Standard Process for Data Mining*), amplamente utilizado em projetos de ciência de dados por oferecer uma estrutura sistemática, iterativa e orientada ao ciclo de vida dos modelos preditivos [7].

Neste estudo, o CRISP-DM não é seguido de forma estritamente sequencial, e como tal, foi definida uma versão adaptada às especificidades da previsão de consumo energético e da deteção de anomalias em séries temporais. Esta adaptação enfatiza o carácter cíclico do

processo, permitindo que os resultados obtidos na fase de avaliação conduzam a revisões na fase de modelação, nomeadamente através do ajuste e exploração de hiperparâmetros, sempre que necessário, para melhorar o desempenho preditivo. A Figura 1 apresenta a adaptação do ciclo CRISP-DM utilizada neste trabalho, em que as fases de compreensão do problema, pré-processamento dos dados, modelação e avaliação são organizadas de forma a refletir o foco na previsão do consumo energético e na utilização operacional das previsões.

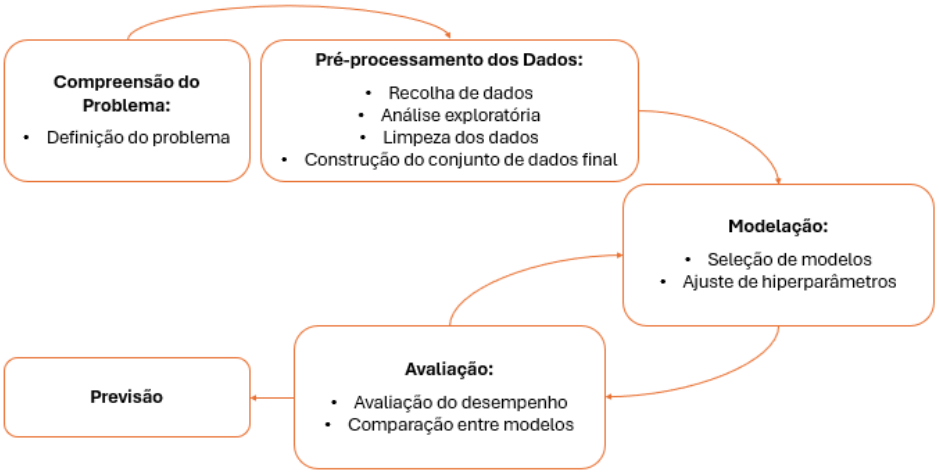

Figura 1: Metodologia adaptada CRISP-DM

A fase de compreensão do problema é dedicada à clarificação dos requisitos de estimação. O pré-processamento dos dados inclui a recolha dos registos históricos de consumo energético, referentes ao período de 2015 a 2023, seguida de uma análise exploratória, permitindo compreender a evolução temporal do consumo. Posteriormente, foram realizadas tarefas de preparação dos dados, nomeadamente a remoção dos dias incompletos existentes no início e no fim do conjunto de dados, de forma a garantir a consistência temporal da série. Os dados foram também normalizados e organizados em formato adequado à modelação, procedendo-se à sua separação em conjuntos de treino, validação e teste. A fase de modelação envolveu a seleção e implementação de diferentes modelos de estimação do consumo energético. Para cada modelo, foi realizado um processo de ajuste de hiperparâmetros. Na fase de avaliação, foi analisado o desempenho dos modelos com base em métricas quantitativas adequadas ao problema. Por fim, na fase de previsão, os modelos treinados foram utilizados para gerar previsões horárias do consumo energético, sendo posteriormente agregadas para obter estimativas do consumo diário, que constituem a base para a análise dos desvios entre os valores estimados e observados.

1.3. Estrutura do trabalho

Este trabalho está organizado em sete capítulos, estruturados de forma a permitir uma compreensão progressiva do tema e dos objetivos desta investigação.

O **Capítulo 2 – Enquadramento Teórico**, explora os conceitos fundamentais que sustentam esta dissertação. Este capítulo aborda as bases teóricas associadas aos modelos preditivos, incluindo abordagens de Estatística e de *Machine Learning*, bem como métricas e metodologias comuns na avaliação do desempenho de modelos.

O **Capítulo 3 – Revisão de Literatura**, analisa os avanços mais recentes na previsão e modelação do consumo energético, bem como na deteção de anomalias em sistemas energéticos. Este capítulo revê a literatura científica relevante, destacando as principais contribuições e identificando as lacunas que justificam a necessidade do presente estudo.

O **Capítulo 4 – Preparação e Exploração de Dados**, apresenta os dados utilizados no estudo, incluindo as suas características principais, fontes e etapas de preparação realizadas. Este capítulo aborda o processo de tratamento e preparação dos dados, nomeadamente a análise exploratória, a remoção de registos incompletos e a organização do conjunto de dados para posterior modelação.

O **Capítulo 5 – Desenvolvimento de Modelos**, detalha o processo de desenvolvimento e implementação dos modelos de estimação do consumo energético, incluindo a definição das arquiteturas, o ajuste de hiperparâmetros e a configuração dos procedimentos de treino e validação.

O **Capítulo 6 – Análise dos Resultados**, apresenta e discute os resultados obtidos com os diferentes modelos desenvolvidos. São analisados os desempenhos preditivos com base em métricas de erro adequadas, comparando-se os resultados entre os diferentes modelos. Este capítulo inclui a análise da robustez e consistência dos modelos, bem como a identificação das abordagens com melhor desempenho.

O **Capítulo 7 – Conclusão**, finaliza o trabalho, sintetizando as principais conclusões do estudo e a sua relevância no contexto da estimação do consumo energético e do seu potencial suporte à deteção de anomalias em edifícios institucionais. São ainda discutidas as principais limitações e desafios associados ao desenvolvimento deste trabalho, bem como propostas para trabalho futuro.

2. Enquadramento Teórico

A previsão do consumo energético constitui uma área de crescente relevância no contexto da sustentabilidade e da gestão eficiente dos recursos [8]. A antecipação de padrões de utilização permite não só a otimização de processos e a redução de desperdícios, mas também o apoio à tomada de decisão operacional [9]. Nesta perspetiva, a previsão pode ser vista como um problema de estimação de valores futuros com base na evolução histórica da série de consumo, podendo incluir variáveis externas [10]. Uma vez que os dados de consumo são registados ao longo do tempo, a sua análise enquadra-se no domínio das séries temporais, cuja estrutura condiciona a modelação e o desempenho dos modelos preditivos [9].

O presente capítulo tem como objetivo apresentar os conceitos teóricos fundamentais que sustentam a metodologia adotada nesta dissertação, incluindo as características das séries temporais e a decomposição das suas principais componentes, bem como uma visão geral dos modelos de aprendizagem computacional [11] e das métricas de avaliação mais utilizadas na previsão do consumo energético [12].

2.1. Séries Temporais

As séries temporais são definidas como conjuntos de pontos de dados observados de forma sequencial ao longo do tempo, tipicamente em intervalos regulares, em que a ordem temporal é fundamental para compreender a interdependência entre os dados [13, 14]. Esta dependência entre observações adjacentes constitui um recurso intrínseco das séries temporais, permitindo que o valor atual seja expresso em função de observações passadas [15].

A Figura 2 apresenta um exemplo ilustrativo de uma série temporal de consumo energético registado ao longo de vários anos [16].

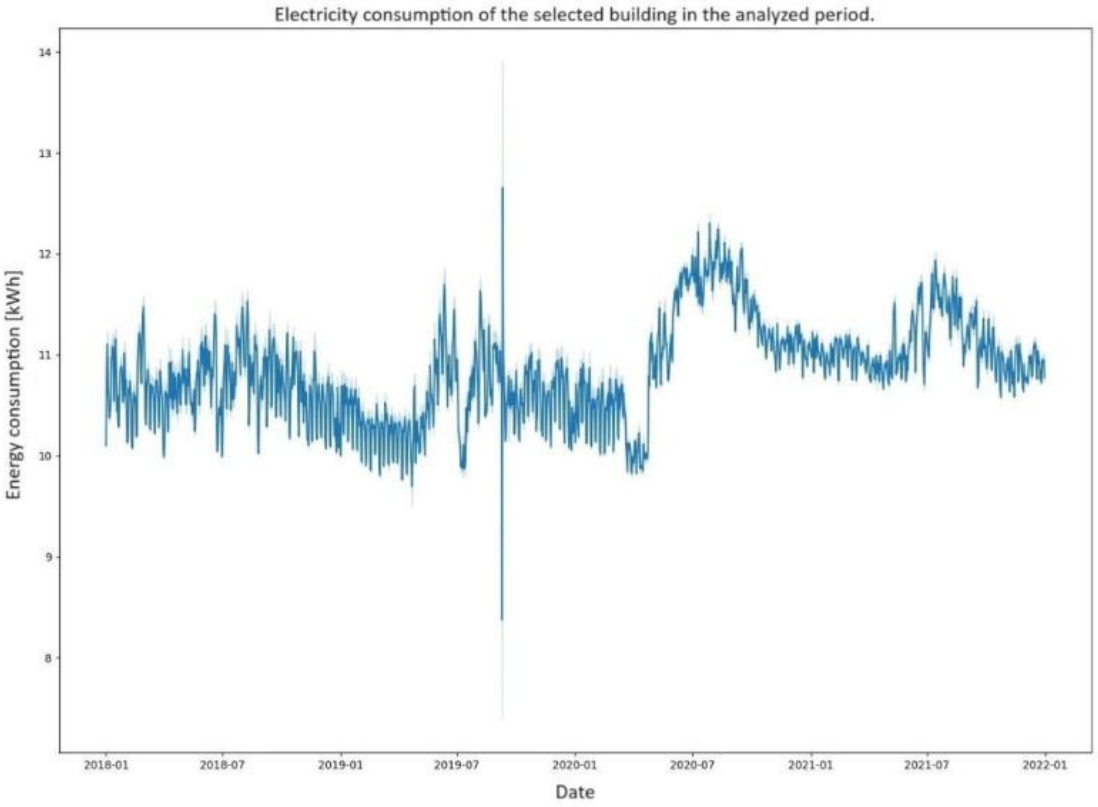

Figura 2: Exemplo de série temporal de consumo energético registado. (Adaptado de [12])

As séries temporais podem ser classificadas como univariadas ou multivariadas, consoante o número de variáveis de saída que se pretende modelar em simultâneo [17]. Uma série temporal univariada foca-se na evolução de uma variável de interesse ao longo do tempo, como, por exemplo, o consumo energético horário de um edifício. Por sua vez, as séries temporais multivariadas envolvem a modelação conjunta de duas ou mais séries temporais distintas, por exemplo, séries de temperaturas registadas simultaneamente em diferentes cidades ou diferentes tipos de carga energética, recorrendo a técnicas específicas que permitem modelar explicitamente as interdependências entre essas séries. Independentemente desta distinção, modelos de previsão podem incorporar variáveis explicativas adicionais (como por exemplo a temperatura ou a humidade), frequentemente denominadas por exógenas, que ajudam a explicar o comportamento das séries principais [18].

A inclusão de variáveis exógenas no estudo de séries temporais tende a melhorar a capacidade preditiva, sobretudo em ambientes onde fatores externos exercem influência significativa sobre o consumo [19].

Outro aspeto relevante das séries diz respeito à estacionaridade. Uma série temporal é considerada estacionária quando as suas propriedades estatísticas, como a média e a variância, se mantêm aproximadamente constantes ao longo do tempo. Em contraste, séries não estacionárias apresentam variações estruturais, sendo comum a necessidade de aplicar transformações para estabilizar a série antes da modelação. A correta identificação do tipo de série e das suas propriedades estatísticas é determinante para a seleção adequada dos modelos de previsão e para a fiabilidade dos resultados obtidos [20].

2.2. Componentes de uma Série Temporal

O comportamento de uma série temporal pode ser interpretado como o resultado da combinação de diferentes componentes estruturais que descrevem a evolução de um fenómeno ao longo do tempo. De forma geral, estas componentes incluem a tendência, a sazonalidade, a componente cíclica e a componente irregular ou aleatória [21, 22].

A componente de **tendência** representa a evolução de longo prazo da série, traduzindo a direção geral do fenómeno ao longo do tempo, podendo manifestar-se sob a forma de crescimento, decréscimo ou estabilidade. A **sazonalidade** corresponde a padrões que se repetem em intervalos regulares e previsíveis, geralmente associados a períodos fixos como horas do dia, dias da semana, meses ou estações do ano. A **componente cíclica** descreve flutuações de médio ou longo prazo que não apresentam uma periodicidade fixa, estando frequentemente associadas a variações estruturais ou mudanças mais amplas no comportamento da série temporal, e muitas vezes, dependentes de fatores externos. Por fim, a **componente irregular** ou **aleatória** engloba variações imprevisíveis não explicadas pelas restantes componentes, sendo geralmente atribuídas a fatores ocasionais, ruído ou eventos externos inesperados [22].

A Figura 3 apresenta um exemplo ilustrativo de uma série temporal observada, evidenciando a evolução do fenómeno ao longo do tempo e a Figura 4 apresenta a decomposição dessa série temporal nas suas principais componentes, evidenciando a separação entre a tendência, a sazonalidade, a componente cíclica e a componente irregular [23].

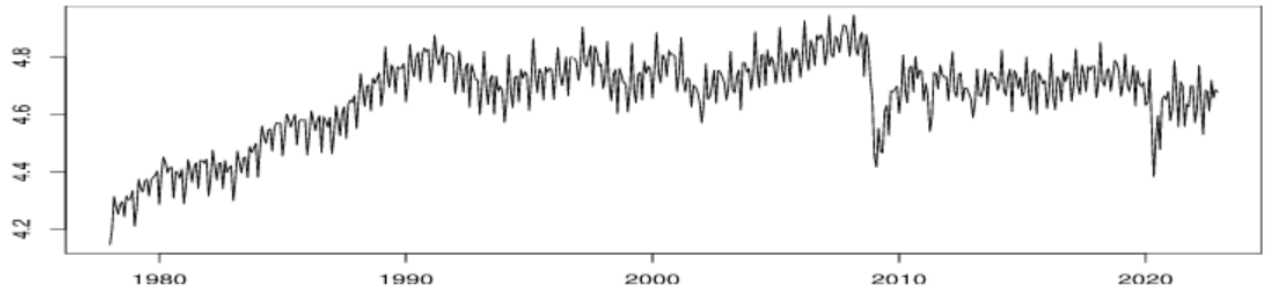

Figura 3: Série temporal observada do índice de produção industrial no Japão (dados de janeiro de 1978 a dezembro de 2022). Adaptado de [23].

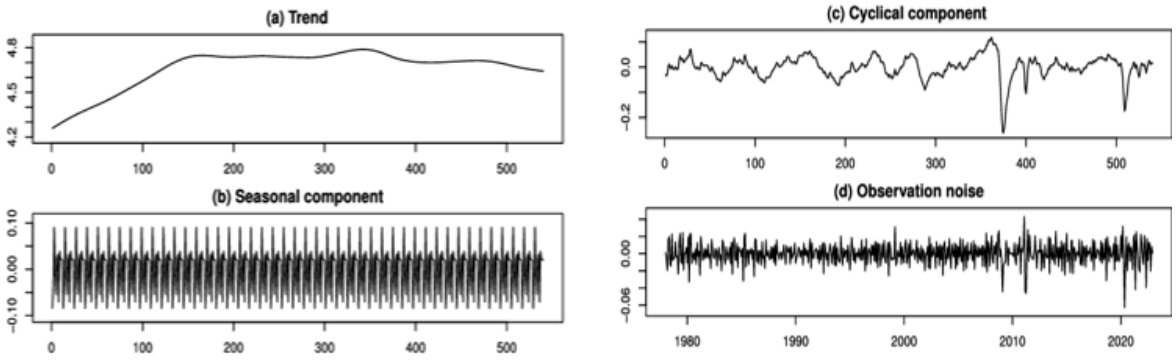

Figura 4: Decomposição da série temporal nas suas componentes, obtido através de um modelo de espaço de estados. Adaptado de [23].

2.3. Modelos de Previsão de Séries Temporais

A previsão de séries temporais constitui um problema central em diversas áreas de aplicação, assumindo particular relevância no contexto energético, onde a antecipação do consumo permite apoiar o planeamento de recursos, a operação de sistemas e a tomada de decisão [24].

A evolução das metodologias de previsão reflete uma transição de modelos estatísticos mais simples para arquiteturas computacionais com maior capacidade de capturar padrões temporais complexos e relações não lineares. Estas abordagens são genericamente agrupadas em duas grandes categorias: os modelos estatísticos clássicos, tradicionalmente baseados em pressupostos de linearidade e estacionariedade, e os modelos baseados em aprendizagem computacional, que utilizam processos iterativos para extrair características intrínsecas diretamente de grandes volumes de dados [25]. Enquanto os métodos estatísticos primam pela interpretabilidade e robustez em séries curtas, os modelos de aprendizagem automática

demonstram uma capacidade superior de generalização perante comportamentos dinâmicos e multivariados característicos de séries temporais mais longas e de maior complexidade [26].

2.3.1. Modelos Estatísticos Clássicos

Os modelos estatísticos clássicos baseiam-se na premissa de que o comportamento futuro de uma série temporal pode ser explicado a partir dos padrões observados nos seus valores passados. De um modo geral, estes métodos pressupõem que a série seja estacionária, ou seja, que apresente a média e a variância praticamente constantes ao longo do tempo. Quando esta condição não é satisfeita, recorre-se frequentemente a transformações, como a diferenciação, a logaritmização ou outras transformações de Box-Cox [27], de forma a tornar a série compatível com os pressupostos do modelo. Entre os modelos estatísticos clássicos mais utilizados na previsão de séries temporais destacam-se os modelos ARIMA (*AutoRegressive Integrated Moving Average*), e a sua extensão sazonal, o SARIMA (*Seasonal AutoRegressive Integrated Moving Average*), amplamente utilizados pela sua capacidade de modelar dependências temporais e, no caso do SARIMA, incorporar padrões sazonais recorrentes [28].

2.3.1.1. ARIMA

O modelo ARIMA estrutura-se através da combinação de três componentes fundamentais, que permitem representar dependências temporais de curto prazo numa série temporal previamente transformada:

- **AR ($\boldsymbol{p}$) – componente autorregressiva**: modela a relação entre a observação atual e p valores passados na série.
- **I ($\boldsymbol{d}$) – componente integrada**: corresponde ao número de diferenciações aplicadas à série com o objetivo de garantir a estacionariedade.
- **MA ($\boldsymbol{q}$) – componente de média móvel**: modela a dependência entre a observação atual e q erros de previsões anteriores.

A combinação destas três componentes é habitualmente representada pela notação $ARIMA(p, d, q)$, na qual os respetivos parâmetros controlam a complexidade e a capacidade de ajuste do modelo. O modelo ARIMA é particularmente adequado para séries temporais sem padrões sazonais explícitos, nas quais a dinâmica temporal é dominada por tendências e variações de curto prazo [17].

2.3.1.2. SARIMA

O modelo SARIMA constitui uma extensão do modelo ARIMA, desenvolvida para lidar com séries temporais que apresentam padrões sazonais regulares. Para além das dependências temporais de curto prazo, este modelo permite capturar relações periódicas associadas à repetição sistemática de comportamentos ao longo do tempo. O SARIMA integra componentes não sazonais, análogas às do ARIMA, e componentes sazonais adicionais, responsáveis por modelar a sazonalidade da série temporal. Estas componentes podem ser descritas da seguinte forma:

- **SAR ($\boldsymbol{P}$) – Componente autorregressiva sazonal:** modela a relação entre observações separadas por múltiplos do período sazonal (s).
- **SI ($\boldsymbol{D}$) – Componente integrada sazonal:** corresponde ao número de diferenciações sazonais aplicadas à série temporal, com o objetivo de remover padrões periódicos persistentes.
- **SMA ($\boldsymbol{Q}$) – Componente de média móvel sazonal:** modela a dependência entre a observação atual e erros de previsão sazonais passados.

A combinação das componentes não sazonais e sazonais é habitualmente representada pela notação $ARIMA(p,d,q)(P,D,Q)_s$, em que s corresponde ao período da sazonalidade. O modelo SARIMA é particularmente adequado para séries temporais com padrões sazonais bem definidos, mantendo os pressupostos da linearidade e estacionariedade características dos modelos clássicos [17].

2.3.2. *Machine Learning*

A evolução das técnicas de análise e previsão de dados conduziu à transição de modelos estatísticos clássicos, baseados em pressupostos bem definidos, para métodos capazes de aprender padrões diretamente através da informação observada. Esta mudança tornou-se particularmente relevante em contextos caracterizados por elevada complexidade, variabilidade e não linearidade, nos quais abordagens tradicionais podem revelar limitações na capacidade de representação dos dados [18].

Neste contexto, a Aprendizagem Computacional (ML, do inglês *Machine Learning*) surge como uma abordagem central, focada no desenvolvimento de modelos que aprendem relações a partir dos dados e generalizam o conhecimento adquirido, permitindo produzir previsões fiáveis quando aplicados a cenários novos e não observados. O desempenho destes

modelos resulta do ajustamento iterativo dos seus parâmetros durante um processo de treino, no qual se procura otimizar uma função objetivo que represente adequadamente as relações subjacentes nos dados.

Um aspeto fundamental neste processo é o compromisso entre capacidade de ajuste e generalização. Modelos excessivamente complexos tendem a memorizar o conjunto de treino, capturando ruído estatístico em detrimento de padrões relevantes, fenómeno designado por *overfitting*. Por este motivo, a avaliação do desempenho em dados não utilizados no treino assume um papel central no desenvolvimento de modelos de ML [29].

De forma geral, os métodos de ML podem ser classificados de acordo com o paradigma de aprendizagem adotado [30]:

- **Aprendizagem supervisionada**, na qual o modelo é treinado a partir de dados rotulados, sendo amplamente utilizada em problemas de regressão e classificação;
- **Aprendizagem não supervisionada**, que procura identificar estruturas latentes ou agrupamentos em dados não rotulados;
- **Aprendizagem por reforço**, baseada na interação contínua com um ambiente, recorrendo a mecanismos de recompensa para otimizar o processo de decisão.

No âmbito da aprendizagem supervisionada, o *Deep Learning* (DL), enquanto subconjunto de métodos baseados em redes neuronais artificiais com múltiplas camadas, tem assumido particular relevância devido à sua capacidade de modelar relações complexas e não lineares. Estas características tornam o DL especialmente adequado para a previsão de séries temporais, nas quais são comuns dependências sequenciais e padrões de elevada variabilidade, como no consumo energético ou em mercados financeiros [31].

Para aplicar a aprendizagem supervisionada a séries temporais, é necessário reformular a série original em pares de entrada–saída através da técnica de janelas deslizantes (*sliding windows*). Neste procedimento, um conjunto fixo de observações passadas é utilizado como entrada para prever valores futuros, permitindo transformar o problema temporal num formato compatível com modelos de ML [32].

A Figura 5 ilustra este processo, evidenciando a separação entre a janela de entrada e o horizonte de previsão [33]. Consoante o objetivo do problema, a previsão pode ser realizada de forma *one-step ahead*, correspondente à estimativa do instante seguinte, ou *multi-step*

ahead, na qual são previstos múltiplos instantes futuros em simultâneo, sendo o horizonte de previsão definido de acordo com os requisitos da aplicação.

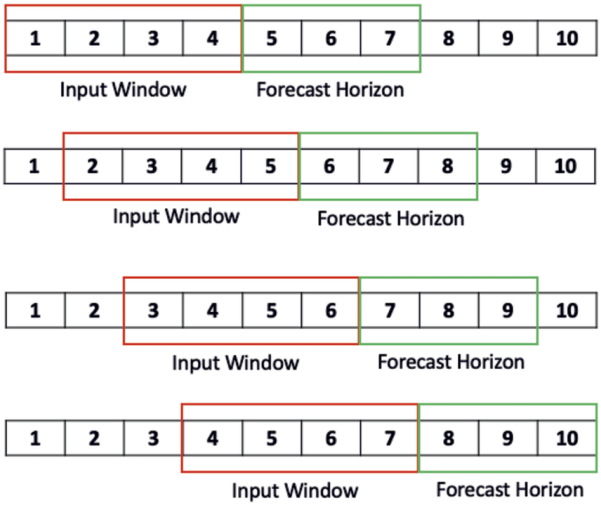

Figura 5: Janelas Deslizantes [33]

No presente estudo, foram considerados seis modelos de ML e DL para a previsão do consumo energético, incluindo arquiteturas baseadas em redes neuronais artificiais e métodos de *ensemble*. A seleção destes modelos teve como objetivo comparar diferentes abordagens de previsão de séries temporais, abrangendo técnicas amplamente utilizadas na literatura. Em seguida, é apresentada uma breve descrição das principais características de cada modelo utilizado neste estudo.

MLP

A *Multilayer Perceptron* (MLP) é uma rede neuronal do tipo *feedforward*, estruturada em camadas sucessivas onde a informação se propaga unidireccionalmente desde a camada de entrada até à camada de saída. A camada de entrada é responsável por receber as variáveis do problema, enquanto as camadas ocultas (ou intermédias) realizam transformações progressivas nos dados, o que permite identificar padrões complexos. Por sua vez, a camada de saída produz a estimativa final associada à tarefa em estudo [31]. A Figura 6 ilustra esta arquitetura, evidenciando a organização hierárquica e o fluxo de informação [34].

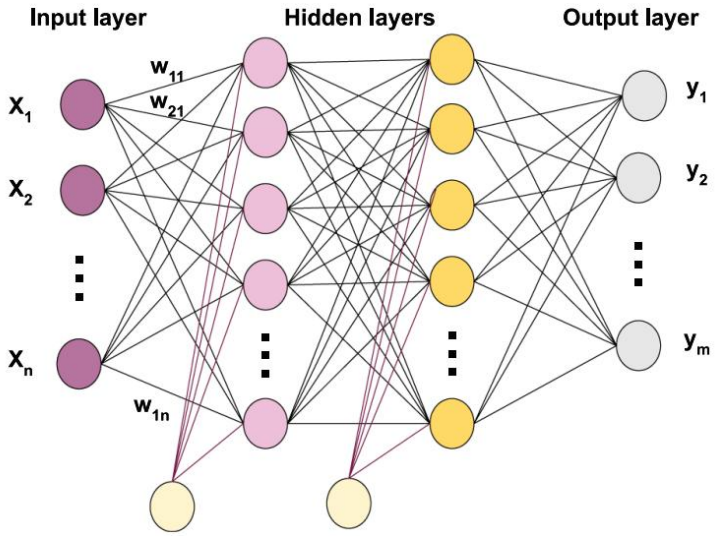

Figura 6: Modelo MLP [34]

Em cada camada, os neurónios combinam os valores de entrada através de pesos, produzindo uma soma ponderada, à qual é aplicada uma função de ativação [35]. Este mecanismo introduz não linearidade no modelo, conferindo à MLP a capacidade de aprender representações que superam as limitações de modelos lineares [36]. A utilização de múltiplas camadas ocultas aumenta a capacidade da MLP em modelar estruturas mais elaboradas dos dados [37].

O processo de aprendizagem assenta no ajuste iterativo dos pesos para minimizar o erro entre as previsões e os valores reais, habitualmente através do algoritmo de retro propagação do erro (*backpropagation*) [36]. Este processo é frequentemente combinado com otimizadores como o Gradiente Descendente Estocástico (SGD) ou o *Adam*, sendo este último um dos métodos mais utilizados e eficazes para otimizações estocásticas [37]. Durante o treino, o erro calculado na saída é propagado inversamente pela rede, permitindo atualizar as ligações de acordo com o contributo de cada neurónio para o desvio total [36].

Para preservar a capacidade de generalização e mitigar o *overfitting*, recorre-se a técnicas de regularização como o *dropout* ou a penalização de pesos [38]. O *overfitting* ocorre quando a rede se ajusta excessivamente aos dados de treino, aprendendo não apenas os padrões relevantes, mas também o ruído e as flutuações aleatórias presentes nesses dados, o que compromete o seu desempenho quando aplicados a dados não observados [31]. As estratégias de regularização contribuem para o controlo da complexidade do modelo e para a melhoria da capacidade de generalização, resultando numa aprendizagem mais estável e robusta [36].

LSTM

A rede *Long Short-Term Memory* (LSTM) é uma arquitetura de Redes Neuronais Recorrentes (RNN) concebida para modelar dados sequenciais e capturar dependências de longo prazo [39]. Ao contrário do fluxo estritamente unidirecional característico do MLP, a LSTM incorpora ligações recorrentes que permitem a propagação e persistência da informação ao longo do tempo [31]. A sua estrutura fundamental assenta numa célula de memória, responsável por armazenar e atualizar informação relevante ao longo de sucessivos instantes temporais, possibilitando a preservação de dependências passadas [40]. A Figura 7 ilustra a estrutura típica de uma unidade LSTM, destacando as suas principais componentes [41].

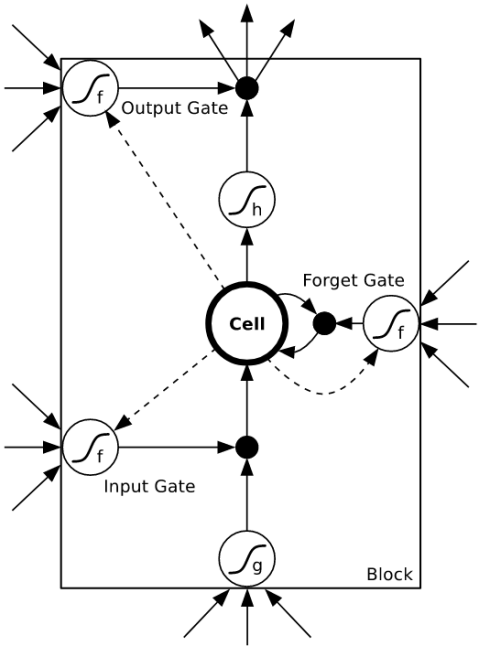

Figura 7: Arquitetura LSTM [41]

O funcionamento da LSTM é regulado por três mecanismos principais designados por portas (*gates*):

- Porta de esquecimento (*Forget gate*): que determina que informação previamente armazenada deve ser descartada;
- Porta de entrada (*Input gate*): é responsável por controlar a incorporação de nova informação na célula de memória;

- Porta de saída (*Output gate*): define a informação transmitida para o estado oculto e para o instante temporal seguinte.

Estas portas operam com funções de ativação que regulam o fluxo de informação, permitindo decidir dinamicamente o que manter ou descartar em cada passo temporal. Este mecanismo de memória é crucial para mitigar o desaparecimento do gradiente, fenómeno em que os sinais de erro se atenuam ao longo do tempo e dificultam a aprendizagem de dependências distantes em redes recorrentes convencionais. Ao permitir que a informação relevante atravesse a célula de memória com menor degradação, a arquitetura LSTM mantém gradientes mais estáveis e a capacidade de modelar relações de longo prazo [41].

O processo de aprendizagem da LSTM assenta no ajuste iterativo dos pesos das portas e das ligações recorrentes através da Retropropagação Através do Tempo (*Backpropagation Through Time* – BPTT), minimizando o erro entre previsões e valores reais [41]. Utilizam-se otimizadores como o *Adam* [37], que ajustam os parâmetros de aprendizagem de forma adaptativa para lidar com a variabilidade dos dados sequenciais e facilitar a convergência em problemas complexos de séries temporais [31].

Tal como no modelo MLP, a rede LSTM recorre frequentemente a técnicas de regularização, nomeadamente *dropout* adaptado a estruturas recorrentes e penalização de pesos, com o objetivo de reduzir o risco de *overfitting* e melhorar a capacidade de generalização do modelo [42].

GRU

A rede *Gated Recurrent Units* (GRU) é uma arquitetura de Redes Neuronais Recorrentes concebida para modelar dependências temporais em dados sequenciais, proposta como uma alternativa mais simplificada à LSTM [43]. À semelhança da LSTM, a GRU incorpora mecanismos de portas que regulam o fluxo de informação ao longo do tempo, mas distinguem-se por possuir uma estrutura mais compacta, com apenas duas portas em vez de três, e sem uma célula de memória separada [44]. Esta simplificação arquitetural resulta num menor número de parâmetros a treinar, o que pode conduzir a uma convergência mais rápida e a uma redução do custo computacional, mantendo uma capacidade de modelação comparável, em muitos contextos, à LSTM. A Figura 8 ilustra a estrutura típica de uma unidade GRU, destacando as suas principais componentes [45].

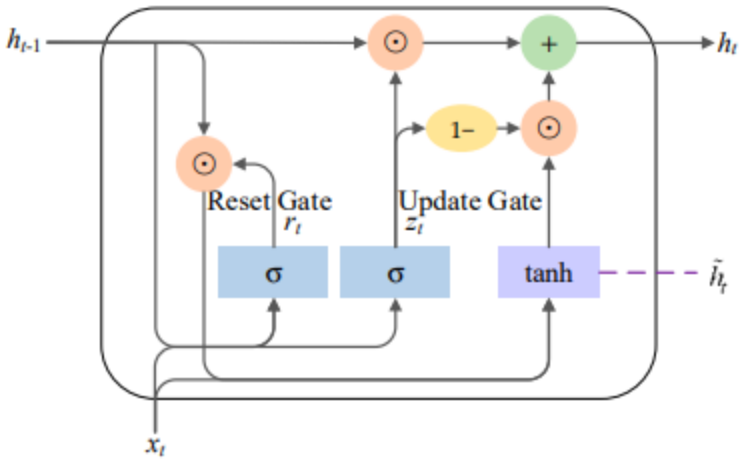

Figura 8: Arquitetura GRU [45]

A estrutura da GRU é regulada por dois mecanismos principais designados por portas (*gates*) [45]:

- Porta de atualização (*Update gate*): que determina quanta da informação anterior deve ser mantida e quanta nova informação deve ser incorporada no estado atual;
- Porta de reinicialização (*Reset gate*): que controla quanta da informação do passado deve ser descartada antes de calcular o novo conteúdo candidato, permitindo que a rede ignore informação que deixara de ser relevante para a sequência.

Ao contrário da LSTM, a GRU não separa a memória do estado oculto, utilizando um único estado para armazenar e atualizar informação relevante ao longo do tempo. O estado atual resulta de uma combinação entre o estado anterior e um novo estado candidato, ponderado pela porta de atualização [46]. Ao fundir os mecanismos de esquecimento e de incorporação de nova informação num único controlo, a GRU consegue regular de forma eficiente a influência do passado em cada novo passo temporal, mantendo dependências de longo prazo com uma arquitetura mais compacta [31].

Tal como outras redes recorrentes, a GRU é treinada por Retropropagação Através do Tempo (BPTT), ajustando iterativamente os pesos das portas e das ligações recorrentes para minimizar o erro entre previsões e valores reais. Podem ser utilizados otimizadores adaptativos, bem como técnicas de regularização, como *dropout*, de forma a melhorar a capacidade de generalização e reduzir o *overfitting* em séries temporais. Devido ao menor

número de parâmetros, a GRU tende ainda a apresentar maior robustez em cenários com conjuntos de dados limitados ou restrições computacionais [43].

N-BEATS

O N-BEATS (*Neural Basis Expansion Analysis for Time Series*) é uma arquitetura de *deep learning* desenvolvida para previsão de séries temporais. O modelo assenta numa abordagem puramente baseada em redes neuronais *feedforward*, sendo constituído exclusivamente por camadas totalmente conectadas, distinguindo-se das arquiteturas recorrentes e convolucionais. Esta opção arquitetural representa uma abordagem alternativa à modelação de dependências temporais, distinguindo-se das arquiteturas baseadas em redes neuronais [47]. Na Figura 9 é apresentado um esquema da arquitetura do N-BEATS.

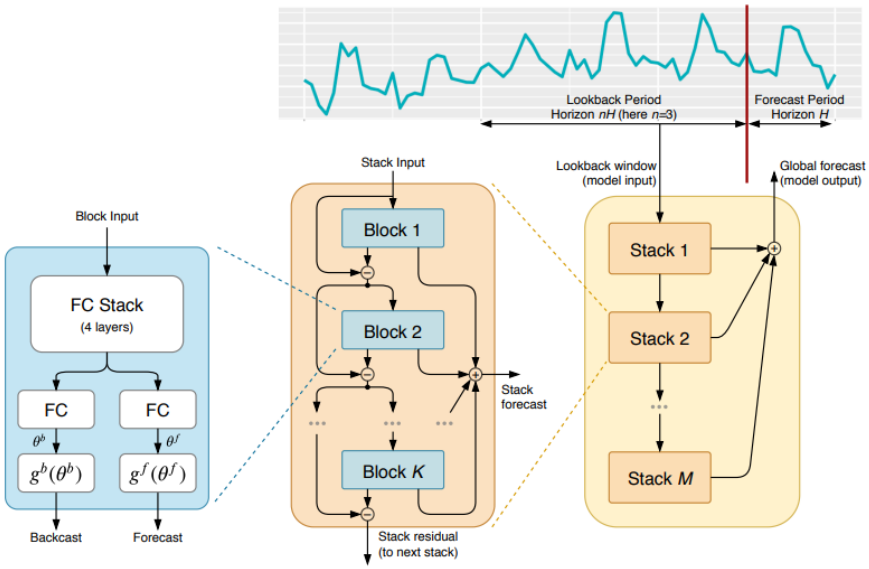

Figura 9: Arquitetura N-BEATS [47]

A arquitetura do N-BEATS é organizada numa sequência de pilhas (*stacks*), sendo que cada uma é composta por múltiplos blocos elementares (*blocks*). Cada bloco recebe como entrada uma janela temporal de observações passadas e produz duas saídas complementares: uma estimativa do passado reconstruído, designado por *backcast*, e uma projeção para o horizonte futuro, denominado por *forecast*. O *backcast* gerado é subtraído da entrada que alimenta o bloco seguinte, originando um mecanismo residual que permite a propagação progressiva da informação já explicada. Este processo favorece uma aprendizagem hierárquica, em que

cada bloco se concentra exclusivamente nos padrões residuais não capturados pelos blocos anteriores [47, 48].

Um elemento central do N-BEATS é a utilização de funções de base (*basis functions*) como forma de representação da série temporal. Em cada bloco, o modelo aprende coeficientes que ponderam essas funções de base, construindo as saídas *backcast* e *forecast* a partir da combinação desses elementos. Na sua configuração interpretável, o N-BEATS é organizado em pilhas funcionalmente distintas: uma pilha de tendência, que utiliza funções de base polinomiais de baixo grau para capturar variações de longo prazo, e uma pilha de sazonalidade, que recorre a séries de *Fourier* para identificar padrões cíclicos e periódicos. Esta decomposição confere ao modelo maior interpretabilidade, permitindo analisar explicitamente a contribuição relativa de cada componente para a previsão final [48].

Do ponto de vista computacional, a ausência de recorrência possibilita um processamento eficiente e paralelizável. O modelo é treinado com otimizadores adaptativos, como o *Adam*, apresentando bom desempenho e aplicabilidade em diferentes contextos de previsão e classificação de séries temporais [48].

N-HiTS

O N-HiTS (*Neural Hierarchical Interpolation for Time Series Forecasting*) é uma arquitetura de *deep learning* desenvolvida para a previsão de séries temporais, proposta como uma extensão do modelo N-BEATS. À semelhança do seu predecessor, o N-HiTS assenta numa abordagem baseada em redes neuronais *feedforward*, sendo constituído por blocos totalmente conectados, sem recorrer a mecanismos recorrentes ou a estruturas especificamente concebidas para a modelação sequencial [49].

A arquitetura do N-HiTS é organizada numa sequência de pilhas (*stacks*), sendo cada pilha composta por múltiplos blocos elementares (*blocks*). Tal como no N-BEATS, cada bloco recebe como entrada uma representação da série temporal e produz componentes de retroprojeção (*backcast*) e projeção (*forecast*), permitindo a remoção progressiva da informação já explicada ao longo da arquitetura [47]. A Figura 10 apresenta um esquema ilustrativo da arquitetura do modelo, evidenciando a sua estrutura hierárquica e o mecanismo de interpolação entre diferentes escalas temporais [49].

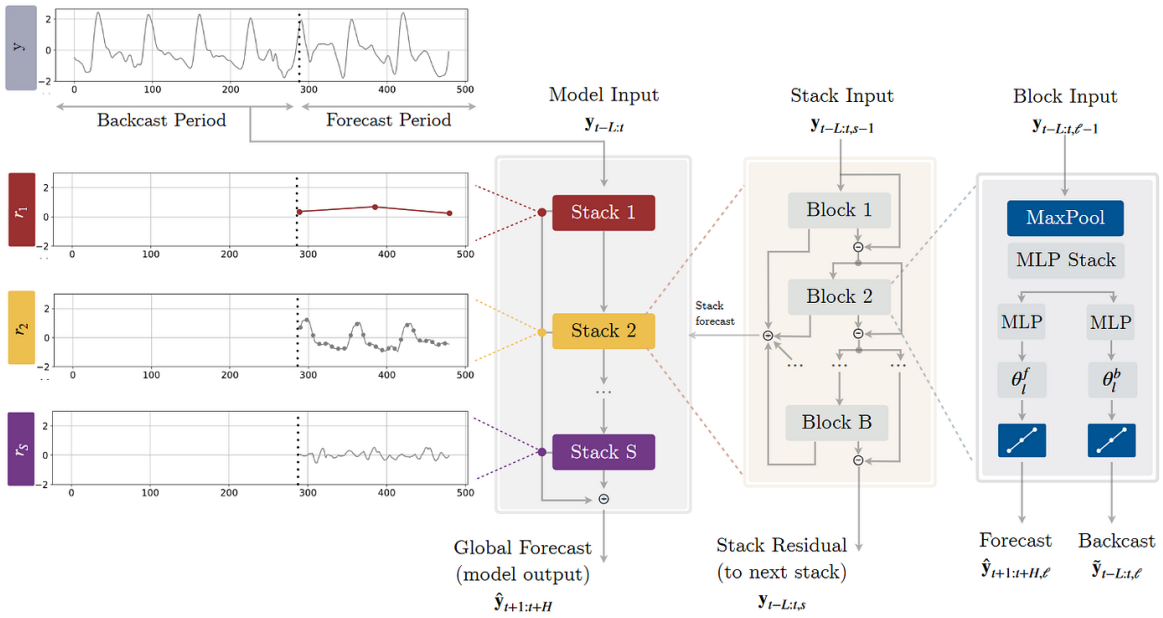

Figura 10: Arquitetura N-HiTs [49]

Uma diferença fundamental relativamente ao N-BEATS reside na forma como o N-HiTS processa a informação temporal ao longo da arquitetura. Em particular, o modelo introduz um mecanismo de amostragem multi-resolução, no qual as entradas dos blocos são obtidas a diferentes resoluções temporais, tipicamente através de operações de *pooling*. Desta forma, os blocos localizados nos níveis iniciais da hierarquia operam sobre representações temporais de menor resolução, capturando padrões de longo prazo, enquanto os blocos subsequentes se concentram progressivamente em componentes de maior detalhe temporal.

Para produzir previsões na escala temporal original, o N-HiTS recorre a mecanismos explícitos de interpolação, que permitem expandir as projeções geradas em resoluções mais baixas para o horizonte futuro completo. A combinação entre amostragem multi-resolução, interpolação hierárquica e aprendizagem residual permite ao modelo reduzir a complexidade associada ao processamento de janelas temporais extensas, mantendo simultaneamente a capacidade de modelar padrões temporais em diferentes escalas [49].

K-Boost

O K-Boost, frequentemente associado a técnicas de *Gradient Boosting*, corresponde a uma família de métodos de aprendizagem baseados na combinação sequencial de múltiplos modelos simples, tipicamente árvores de decisão, com o objetivo de construir um preditor mais robusto. Ao contrário das abordagens baseadas em redes neuronais, o K-Boost assenta

num processo iterativo de melhoria progressiva do modelo, no qual cada novo componente é treinado para corrigir os erros cometidos pelos modelos anteriores [50].

O princípio fundamental do *boosting* consiste na construção de um conjunto de modelos fracos (*weak learners*), que individualmente apresentam desempenho limitado, mas que, quando combinados de forma adequada, resultam num modelo com elevada capacidade preditiva. Em cada iteração, o algoritmo ajusta um novo modelo aos resíduos do modelo acumulado até esse momento, orientando o processo de aprendizagem pela minimização de uma função de perda [50]. Este processo de aprendizagem aditiva permite que o modelo seja construído de forma sequencial através da soma das previsões produzidas por cada árvore individual, formando um modelo de *ensemble*. A previsão final é obtida através da combinação das previsões produzidas por múltiplas árvores de decisão, permitindo aumentar a precisão e a capacidade de generalização do modelo [51].

A Figura 11 apresenta uma representação esquemática do modelo K-Boost, ilustrando o processo de aprendizagem sequencial baseado na construção de múltiplas árvores de decisão, cuja combinação permite melhorar progressivamente o desempenho preditivo do modelo.

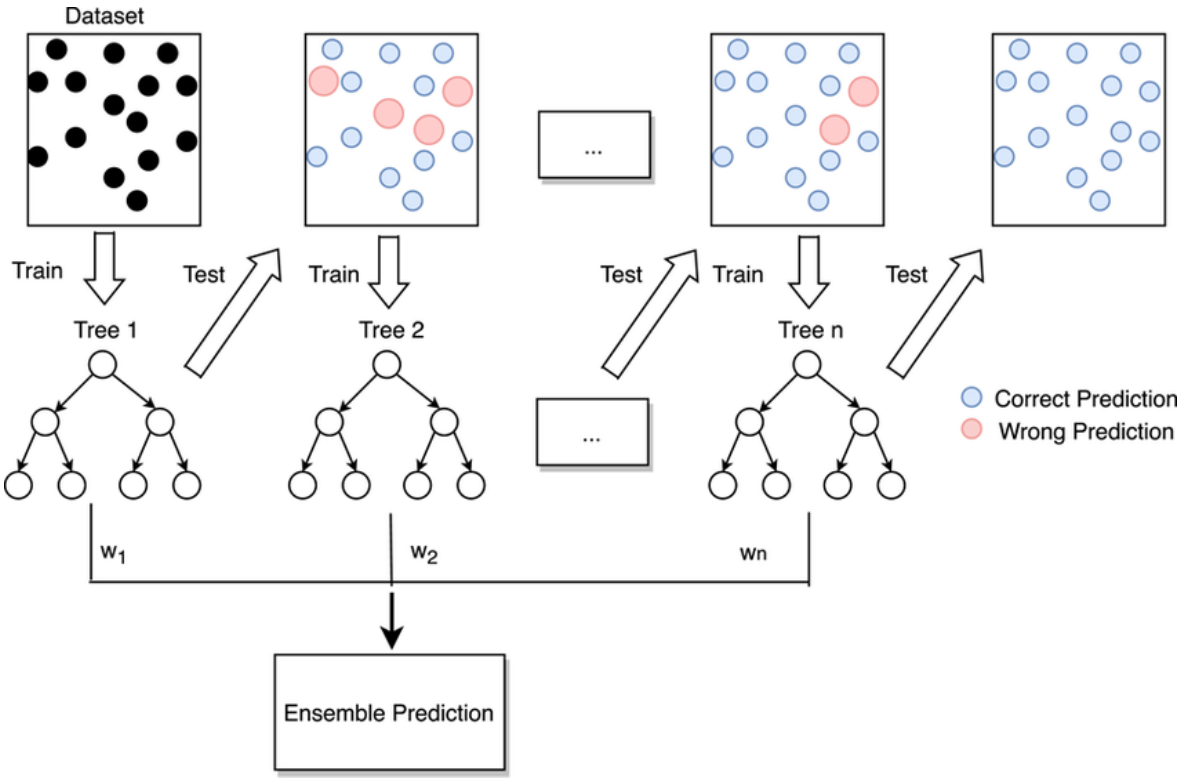

Figura 11: Arquitetura K-Boost [51]

Uma das principais vantagens do K-Boost reside na sua capacidade de modelar relações complexas e não lineares, mantendo simultaneamente um elevado grau de controlo sobre o processo de aprendizagem [51]. A utilização de regularização explícita e estratégias de controlo da complexidade contribui para reduzir o risco de *overfitting*, tornando este tipo de abordagem particularmente eficaz em conjuntos de dados heterogéneos [52].

Devido à sua natureza modular e à eficiência computacional, os métodos baseados em *Gradient Boosting* são amplamente utilizados em problemas de regressão e classificação, sendo frequentemente adotados como modelos de referência ou de comparação face a arquiteturas neuronais mais complexas [53].

2.4. Métricas de Avaliação

A avaliação do desempenho dos modelos preditivos é uma etapa fundamental em qualquer estudo de ML e também em abordagens clássicas de séries temporais baseadas em métodos estatísticos, uma vez que permite quantificar a qualidade das previsões e comparar diferentes abordagens de forma objetiva. Neste contexto, serão apresentadas de seguida algumas das métricas de avaliação mais importantes e mais frequentemente utilizadas na literatura [54], [55].

2.4.1. Erro Absoluto Médio

O Erro Absoluto Médio (MAE, do inglês *Mean Absolute Error*) é definido como

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|,$$

onde y_i representa o valor real observado na i-ésima observação, $\hat{y}_i$ o valor estimado pelo modelo para essa mesma observação e n o número total de observações.
O MAE corresponde à média das diferenças absolutas entre os valores reais e os valores previstos, fornecendo uma medida direta do erro médio cometido pelo modelo, expressa na mesma unidade da variável em estudo.

2.4.2. Erro Quadrático Médio

O Erro Quadrático Médio (MSE, do inglês *Mean Squared Error*) é definido pela expressão

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

Nesta métrica, as diferenças entre os valores reais e previstos são elevadas ao quadrado antes de serem agregadas. Como consequência, erros de maior magnitude têm um impacto significativamente superior no valor final da métrica, tornando o MSE particularmente sensível a previsões muito afastadas dos valores reais.

2.4.3. Raiz do Erro Quadrático Médio

A métrica Raiz do Erro Quadrático Médio (RMSE, do inglês *Root Mean Square Error*) é definida pela expressão

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

que resulta da aplicação da raiz quadrada ao MSE, permitindo que o erro médio seja expresso na mesma escala dos dados originais. Esta métrica combina a penalização de grandes erros com uma interpretação mais intuitiva do erro médio, sendo amplamente utilizada na avaliação de modelos preditivos.

2.4.4. Erro Percentual Médio

O Erro Percentual Médio (MAPE, do inglês *Mean Absolute Percentage Error*) é definido pela expressão

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|.$$

O MAPE mede o erro médio em termos percentuais, calculando a diferença relativa entre os valores previstos e os valores reais. Esta métrica facilita a comparação do desempenho dos modelos em diferentes escalas, embora possa apresentar limitações quando os valores reais y_i assumem valores muito próximos de zero.

2.4.5. Coeficiente de Determinação

O Coeficiente de Determinação (R^2) é dado pela fórmula

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

onde $\bar{y}$ corresponde à média dos valores reais observados. O coeficiente de determinação avalia a proporção da variabilidade dos dados explicada pelo modelo, fornecendo uma medida global da qualidade do ajustamento. Apesar da sua utilidade, o R^2 deve ser analisado em conjunto com métricas de erro absoluto para uma avaliação mais completa do desempenho preditivo.

3. Revisão de Literatura

A recente instabilidade verificada na rede elétrica nacional, evidenciada pelo apagão ocorrido na Península Ibérica em 2025 reforçou a perceção da forte dependência da sociedade contemporânea em relação à energia elétrica. A interrupção do fornecimento, ainda que de forma temporária, revelou o impacto transversal da energia no funcionamento de serviços essenciais, infraestruturas e atividades económicas. Neste contexto, a previsão do consumo energético assume particular relevância, não apenas como instrumento de planeamento e antecipação de necessidades futuras, mas também como referência para avaliar o desempenho real dos sistemas e detetar eventuais desvios operacionais. No presente trabalho, a análise centra-se especificamente em sistemas energéticos associados a *campus* universitários, pelo que a revisão da literatura apresentada neste capítulo privilegia estudos cuja área de aplicação corresponde a este tipo de infraestruturas.

Os dados de consumo energético são recolhidos de forma sequencial ao longo do tempo e, como tal, a sua análise insere-se no domínio das séries temporais. Ao longo dos anos, as abordagens para modelar estes dados foram evoluindo, passando de métodos tradicionais para técnicas de *Machine Learning* e, mais recentemente, para arquiteturas de *Deep Learning*.

Uma vez que a abordagem de deteção de anomalias proposta neste trabalho foca-se em modelos de previsão do consumo energético, o presente capítulo encontra-se organizado em duas secções complementares. Numa primeira fase, são analisadas as abordagens existentes para a previsão do consumo energético em edifícios institucionais, com particular foco em estudos realizados em *campus* universitários. De seguida, são discutidos os métodos de deteção de anomalias em séries temporais energéticas, com ênfase em abordagens baseadas em modelos preditivos, estratégia que fundamenta a metodologia adotada nesta dissertação.

3.1. Previsão do Consumo Energético

A previsão do consumo energético tem sido amplamente estudada nas últimas décadas, resultando no desenvolvimento de uma grande variedade de modelos baseados em diferentes metodologias e técnicas de previsão. As primeiras abordagens estruturadas basearam-se em modelos estatísticos clássicos de séries temporais, particularmente na metodologia de *Box-Jenkins* [27], que originou modelos autorregressivos integrados de médias móveis (ARIMA)

e as suas extensões sazonais (SARIMA). Estes modelos, assentes na modelação de dependências lineares entre as observações passadas e presentes, capturam padrões de tendência e sazonalidade bem definidos, tendo sido amplamente aplicados na previsão do consumo energético ao longo de várias décadas.

No contexto específico de *campus* universitários, os modelos ARIMA e SARIMA têm sido amplamente utilizados para modelar este tipo de dados. No trabalho de Zancanaro et al. [56], por exemplo, foram analisados dados mensais de consumo energético ao longo de 104 meses, tendo como objetivo a previsão do consumo mensal futuro. Seguindo a abordagem clássica de identificação e validação de modelos, os autores selecionaram um modelo SARIMA com diferenciação regular e sazonal anual, adequado à estrutura temporal da série. No período de teste, correspondente a um horizonte de previsão de 24 meses, o modelo apresentou um MAPE de 8,96% e um RMSE de 132kWh.

Para além das análises assentes em dados mensais, diversos trabalhos têm considerado séries temporais com granularidade horária em ambientes universitários. No trabalho de Kim et al. [57] foi analisado o consumo horário de um edifício institucional inserido num *campus* universitário em Seul, utilizando um conjunto de 8.760 observações correspondentes a um ano completo de medições em intervalos de 1 hora. Os autores aplicaram modelos ARIMA e as suas extensões, recorrendo a uma abordagem de *moving window* para avaliar previsões de 1 a 24 passos à frente. Para o horizonte de 1 hora, o modelo ARIMA apresentou um RMSE de 138,69 e um MAPE de 2,28%, enquanto o modelo Reg-ARIMA obteve um RMSE de 138,11 e um MAPE de 2,27%. Para estimar um horizonte de 24 horas, o modelo ARIMA registou um RMSE de 760,42 e um MAPE de 12,01%, verificando-se uma redução da precisão à medida que o horizonte de previsão aumenta.

A crescente variabilidade dos perfis de consumo energético, marcada por múltiplas sazonalidades e forte dependência de fatores externos como as condições meteorológicas, tem vindo a expor limitações dos modelos estatísticos tradicionais, especialmente em cenários de elevada complexidade. Perante estas restrições, a literatura tem vindo a explorar abordagens em *Machine Learning* e *Deep Learning*, capazes de capturar relações não lineares e dependências temporais mais sofisticadas [58].

A introdução de técnicas de ML permitiu alargar o enquadramento metodológico da previsão do consumo energético, possibilitando a modelação de relações não lineares e a integração de múltiplas variáveis explicativas [59]. Para além da estrutura autorregressiva da série

temporal, estes modelos exploram padrões complexos presentes nos dados históricos e em fatores externos, incluindo, por exemplo, variáveis meteorológicas [60].

Estes modelos têm também sido explorados no contexto de edifícios educacionais. No *campus* universitário de Cartagena (Espanha), Ruiz et al. [61] desenvolveu um modelo de previsão recorrendo a métodos *ensemble* baseados em árvores de decisão. O conjunto de dados incluiu medições horárias do consumo energético entre 2011 e 2016, sendo utilizados como preditores variáveis temporais (hora do dia, dia da semana, mês), indicadores de dias especiais e variáveis meteorológicas, bem como valores históricos da carga elétrica. Para um horizonte de previsão de 48 horas, e considerando o ano de 2016 como período de teste, o modelo *Random Forest* apresentou um RMSE de 25,45 kWh e um MAPE de 9,33%, enquanto o modelo *XGBoost* (*eXtreme Gradient Boosting*) obteve um desempenho ligeiramente superior, com RMSE de 23,65 kWh e MAPE de 8,99%. Os resultados evidenciam ainda uma melhoria significativa face a modelos tradicionais, como a Regressão Linear Múltipla (MLR) (MAPE de 19,33%) e o modelo naïve (MAPE de 15,53%), demostrando a eficácia das abordagens baseadas em ML.

Um estudo mais recente foi desenvolvido no *campus* da Universidade de Missouri (EUA) [62], tendo como objetivo a previsão horária do consumo energético, com base nos registos da central de cogeração que o abastece. O conjunto de dados integrou seis anos de medições horárias, entre 2017 e 2022, que inclui tanto o consumo energético do *campus*, como também variáveis meteorológicas como a temperatura do ar, humidade relativa, velocidade e direção do vento, pressão atmosférica e intensidade solar. Foram avaliados diversos algoritmos de ML, nomeadamente *Decision Tree* (DT), *Random Forest* (RF), *Support Vector Regression* (SVR), *K-Nearest Neighbors* (KNN) e XGBoost, sendo os modelos treinados com cinco anos de dados e testados no ano mais recente, recorrendo à validação cruzada temporal em cinco *folds*. No conjunto de teste, o modelo *XGBoost* destacou-se como o modelo mais eficaz, apresentando um RMSE de 1,2078 e MAPE de 3,12%.

Para além dos métodos *ensemble* baseados em árvores de decisão e das máquinas de vetores de suporte, outras abordagens de ML, como as redes neuronais artificiais *feedforward*, têm igualmente sido exploradas no contexto de *campus* universitários. No trabalho de Massana et al. [63] analisou-se a previsão horária do consumo energético em edifícios da Universidade de Girona, recorrendo a diferentes modelos de aprendizagem supervisionada, incluindo MLR, redes neuronais do tipo MLP e SVR. O conjunto de dados integrou

medições horárias com indicadores artificiais de ocupação e variáveis temporais. O estudo centrou-se em dois edifícios (PI e PII). O modelo MLP apresentou um MAPE de 21,31% para o edifício PI e de 34,96% para o edifício PII, enquanto o modelo SVR demonstrou o melhor desempenho em ambos os edifícios, com valores de MAPE de 18,11% no edifício PI e 18,05% no edifício PII.

De forma complementar, Elhabyb et al. [64] analisou a previsão do consumo energético em três edifícios educacionais, utilizando dados recolhidos entre janeiro de 2020 e janeiro de 2023. O conjunto de dados integrou medições horárias do consumo energético, bem como variáveis operacionais dos edifícios e variáveis temporais relevantes para a modelação do consumo. Foram avaliados três algoritmos: RF, LSTM e *Gradient Boosting Regressor* (GBR). Os resultados evidenciaram o desempenho superior do modelo GBR nos três edifícios analisados: CLAS, NHAI e Cronkite. Em termos de MAPE, o modelo alcançou valores de 9,337% para o edifício CLAS, 12,338% para o NHAI e 4,045% para o Cronkite.

Embora os resultados evidenciem que modelos baseados em *boosting* continuam a apresentar elevado desempenho, a crescente complexidade das séries temporais energéticas tem motivado a adoção de arquiteturas de *Deep Learning* especificamente concebidas para modelar dependências temporais de longo alcance. Rene et al. [65] propôs um modelo de previsão conjunta de múltiplas cargas elétricas (arrefecimento, aquecimento e eletricidade), aplicado a dados reais de um *campus* universitário. O estudo integrou dados históricos de consumo energético, variáveis meteorológicas e informação temporal relevante, tendo sido realizada uma análise preliminar das relações entre as diferentes cargas e os fatores externos. Com base nesses dados, foram desenvolvidas três arquiteturas recorrentes: um modelo LSTM padrão, uma versão empilhada (*Stacked* LSTM) com múltiplas camadas recorrentes e uma arquitetura *Bidirectional* LSTM, capaz de explorar dependências temporais em ambas as direções da sequência. O conjunto de dados foi dividido em 90% para treino e 10% para validação e teste. Os resultados evidenciaram que a arquitetura *Stacked* LSTM apresentou o melhor desempenho global. Para a previsão de carga de arrefecimento de verão, o modelo alcançou um MAPE de 2,46%. Verificou-se ainda que a consideração das interdependências entre diferentes tipos de carga permitiu reduzir significativamente o erro de previsão.

Oyinlola e Oluseyi [66] desenvolveram um *framework* de previsão e gestão de carga aplicado ao *campus* da Universidade de Lagos (Nigéria), integrando técnicas de *clustering* e modelos de DL para apoio à sua gestão energética. O conjunto de dados integrou 3.648

registos horários de consumo energético provenientes de 55 edifícios do *campus*, sendo posteriormente agrupados em três *clusters* de procura distintos. Os dados foram divididos em 80% para treino e 20% para teste. Os resultados indicaram que os modelos recorrentes LSTM e GRU apresentaram desempenho inferior face ao modelo *Prophet*, registando um RMSE médio de 48,47 e um MAPE de 54,76%. Os autores atribuem este comportamento à limitação da amostra e ao ruído agregado dos dados.

Apesar do desempenho satisfatório das redes recorrentes, como LSTM e GRU, estas arquiteturas apresentam limitações associadas à sua natureza sequencial, que exige o processamento sucessivo da informação ao longo da série temporal, podendo tornar o treino mais exigente do ponto de vista computacional e dificultar a modelação eficiente de dependências de muito longo alcance [67]. Neste contexto, surgiram arquiteturas concebidas para tarefas de *forecasting*, entre as quais se destacam o N-BEATS [47] e o N-HiTS [49], baseadas em blocos totalmente conectados e estruturas hierárquicas de decomposição temporal.

Apesar de a literatura específica sobre a aplicação destes modelos em previsões de consumo energético em *campus* universitários ser ainda limitada, existem alguns artigos que os utilizam na previsão de consumos energéticos em edifícios. No trabalho de Shaikh et al. [68], analisou-se a previsão de curto prazo do consumo energético num contexto de *smart grid*, utilizando os dados reais de 169 clientes residenciais de Londres, correspondentes a aproximadamente 29 meses de medições convertidas para o consumo diário. O conjunto de dados foi dividido em 75% para treino (126 clientes) e 25% para teste (43 clientes). O modelo N-BEATS foi comparado com diferentes arquiteturas de DL, nomeadamente LSTM, GRU, Blocked LSTM, Blocked GRU e Temporal Convolutional Network (TCN). Para o cenário de um dia à frente (*day-ahead*), o modelo N-BEATS com suavização dos dados e sem variáveis explicativas apresentou um MAPE de 47,04% e um RMSE de 2,27, superando as variantes Blocked LSTM (MAPE 48,78% e RMSE 2,41) e Blocked GRU (MAPE 49,72% e RMSE 2,42). No horizonte de 7 dias, o N-BEATS com dados suavizados obteve um MAPE de 50,69% e RMSE de 2,51, mantendo desempenho superior às arquiteturas recorrentes comparadas. Para o horizonte de 30 dias, o modelo registou um MAPE de 58,79% e um RMSE de 2,87, evidenciando maior robustez na modelação de padrões multi-escala quando comparado com LSTM, GRU e TCN.

No trabalho de Maragkos e Refanidis [18], analisou-se a previsão do consumo energético com base em dados reais de carga elétrica provenientes de 24 países europeus, complementados com variáveis meteorológicas associadas. O conjunto de dados integrou diferentes níveis de agregação temporal (horária, diária e mensal) sendo considerados horizontes de curto prazo (24 horas, 2 dias ou 2 meses, dependendo da agregação) e de longo prazo (168 horas, 365 dias ou 12 meses). Relativamente ao modelo N-HiTS, os resultados evidenciaram um desempenho particularmente relevante em cenários de longo prazo. Na agregação diária com horizonte de longo prazo, o modelo N-HiTS univariado apresentou o menor MAPE em 41,7% dos casos analisados, enquanto a versão multivariada obteve o melhor desempenho em 20,8% dos casos. Em cenários de agregação horária e longo prazo, o N-HiTS destacou-se igualmente, figurando entre os modelos com maior frequência de menor MAPE.

Em síntese, a literatura evidencia que a eficácia da previsão do consumo energético depende fortemente da natureza dos dados, da granularidade temporal e do horizonte de previsão considerado, não existindo uma abordagem universalmente superior para todos os contextos.

3.2. Deteção de anomalias

Segundo Chandola et al. [69], a deteção de anomalias, também designada por *outlier detection* ou *novelty detection*, corresponde ao processo de identificação de padrões nos dados que não estão conforme o comportamento esperado. Estes desvios podem resultar de erros de medição, falhas de sensores, comportamentos fraudulentos ou alterações reais no estado do sistema. Aggarwal [70] define anomalias como instâncias cuja geração não pode ser explicada pelo modelo implícito que descreve o comportamento normal dos dados.

A literatura distingue diferentes tipos de anomalias dependendo da natureza do desvio observado. Uma das classificações mais amplamente utilizadas é apresentada por Chandola [69], que identifica três categorias principais:

- **Anomalias Pontuais (*Point Anomalies*):** correspondem a observações individuais que se desviam significativamente do restante conjunto de dados. Num contexto energético, um pico súbito de consumo elétrico muito superior ao padrão habitual pode ser considerado uma anomalia pontual;
- **Anomalias Contextuais (*Contextual Anomalies*):** ocorrem quando uma observação é anómala apenas num determinado contexto específico. Por exemplo, um consumo

elevado às 03h00 pode ser considerado anómalo, enquanto o mesmo valor durante o horário laboral pode ser perfeitamente expectável;

- **Anomalias Coletivas (*Collective Anomalies*):** verificam-se quando um conjunto de observações consecutivas, analisadas em conjunto, apresentam um comportamento anómalo, ainda que cada ponto individualmente não aparente estar fora do padrão. Em série temporais, este tipo de anomalias é particularmente relevante, podendo refletir alterações estruturais no sistema, como falhas progressivas de equipamentos ou mudanças operacionais.

Diversos estudos na literatura têm explorado métodos baseados em aprendizagem computacional para identificar padrões anómalos em séries temporais. De forma geral, as abordagens procuram aprender o comportamento normal do sistema e identificar desvios significativos relativamente a esse padrão. Um exemplo é apresentado por Yang et al. [71] que propôs um método para deteção de anomalias em redes elétricas inteligentes baseado num modelo híbrido que combina um *Autoencoder* Convolucional, do *inglês Convolutional Autoencoder* (CAE), com uma rede GRU. O conjunto de dados utilizado foi recolhido da *China Shouthern Power Grid*, contendo medições de consumo elétrico de utilizadores individuais e leituras agregadas de contadores principais em diferentes zonas de fornecimento ao longo de um período de um ano. Neste trabalho, os dados são processados através de *sliding windows* para capturar dependências temporais. O modelo combina um *Convolutional Autoencoder* (CAE) para extração automática de características com uma rede *Gated Recurrent Unit* (GRU) responsável por modelar as dependências da série. A deteção de anomalias é realizada com base no erro de reconstrução, sendo que os valores cuja diferença ultrapassa um determinado limiar são classificados como anómalos. A avaliação do método é realizada através de métricas clássicas de classificação, nomeadamente *accuracy*, *precision*, *recall* e *F1-score*, demonstrando melhor desempenho quando comparado com modelos CAE, CNN e GRU aplicados individualmente.

Ravinder e Kulkarni [72] propõem um método de deteção de anomalias em *smart grids* baseada numa arquitetura recorrente com unidades *Dilated* GRU. O conjunto de dados utilizado inclui medições provenientes de *smart meters*, dados meteorológicos e padrões de comportamentos de consumidores, permitindo modelar múltiplos fatores que influenciam o consumo energético. Antes da aplicação do modelo de deteção, é aplicado um método de seleção de características denominado *Modified Flow Direction Algorithm* (MFDA), que identifica os atributos mais relevantes para a análise. As características selecionadas são

posteriormente introduzidas numa arquitetura denominada *Adpative Residual Recurrent Neural Network*, responsável por modelar dependências temporais complexas. As anomalias são identificadas como desvios relativamente aos padrões aprendidos durante o treino, sendo avaliadas através de métricas como taxa de deteção, precisão e taxa de falsos positivos, demonstrando melhorias relativamente a métodos tradicionais.

Sorensen et al. [73] apresenta uma abordagem não supervisionada para deteção de anomalias em séries temporais multivariadas industriais. O estudo baseia-se na utilização de modelos de previsão temporal com o objetivo de identificar desvios relativamente ao comportamento normal de processos industriais. Foram utilizados dados reais recolhidos de ferramentas de fabrico industrial correspondentes a 912 execuções de processos, contendo medições de sensores tais como pressão, temperatura e outras variáveis operacionais registadas em intervalos de 0.1 segundos. A metodologia proposta baseia-se numa abordagem de previsão temporal, na qual um conjunto de observações passadas é utilizado como entrada para prever um conjunto de observações futuras. Inicialmente, o modelo é treinado para prever valores futuros da série temporal utilizando dados históricos considerados predominantemente normais. Durante a fase de teste, os valores previstos são comparados com os valores reais observados, sendo identificadas anomalias quando o erro de previsão ultrapassa um limiar previamente definido. No estudo são avaliados dois modelos de previsão, nomeadamente o N-BEATS e as *Graph Neural Networks* (GNN). O desempenho do modelo é avaliado considerando diferentes combinações destas janelas temporais, permitindo analisar o impacto da quantidade de informação histórica utilizada na previsão na capacidade de deteção de anomalias. A deteção de anomalias é realizada através da comparação entre os valores previstos pelo modelo e os valores reais observados na série temporal. Para cada instante temporal, o modelo gera uma previsão com base nas observações anteriores, sendo posteriormente calculada a diferença entre o valor previsto e o valor efetivamente observado. Essa discrepância é quantificada através do MSE. A partir destes erros é calculado um *score* de anomalia, obtido com base nos maiores erros registados entre as variáveis monitorizadas, sendo um comportamento considerado anómalo quando esse *score* ultrapassa um limiar previamente definido. No estudo considerado, foi utilizado um limiar fixo de 0.1, definido empiricamente pelos autores e mantido constante ao longo das experiências realizadas. Importa salientar que, na ausência de dados rotulados contendo anomalias reais, os autores recorrem à introdução de anomalias sintéticas nos dados de teste, permitindo avaliar de forma controlada a capacidade dos modelos em identificar diferentes tipos de desvios

relativamente ao comportamento normal do sistema. Foram consideradas três categorias principais de anomalias: alterações de amplitude (*amplitude shift*), correspondentes a mudanças inesperadas no valor do sinal; deslocamentos temporais (*time shift*), associados a atrasos ou antecipações na ocorrência de eventos; e alterações na duração de estados (*step shift*), resultantes de modificações no tempo de permanência de um determinado estado operacional. Estas diferentes categorias de anomalias permitem simular situações plausíveis em contextos industriais e avaliar a robustez dos modelos em cenários com diferentes características de comportamentos anómalos. Os resultados são apresentados predominantemente sob a forma de gráficos, evidenciando a evolução do desempenho dos modelos em função do horizonte de previsão, não sendo fornecidos valores numéricos exatos para todas as métricas consideradas. Ainda assim, verifica-se que o modelo baseado em GNN apresenta melhor desempenho global na deteção de anomalias em séries temporais multivariadas.

Em suma, a literatura apresenta avanços significativos na aplicação de métodos baseados em dados para a identificação de comportamentos anómalos em sistemas energéticos. No entanto, muitos destes estudos focam-se em contextos como ambientes industriais ou redes de distribuição. No caso de edifícios institucionais, como *campus* universitários, o consumo energético é fortemente influenciado por fatores humanos e pela variabilidade associada aos calendários académicos. Neste contexto, torna-se pertinente investigar abordagens baseadas em modelos de previsão do consumo energético, nos quais a análise de desvios entre valores previstos e observados possa funcionar como um mecanismo de monitorização adaptado à realidade destes edifícios.

4. Preparação e Exploração de Dados

Este capítulo aborda os detalhes do conjunto de dados utilizados neste estudo, organizados em seções que descrevem a sua origem, formato, e os procedimentos aplicados para o seu tratamento.

4.1. Descrição dos Dados

Os dados utilizados neste estudo, representam o consumo energético do *campus* 2 do Instituto Politécnico de Leiria, em Portugal, entre 27 de outubro de 2015 e 14 de dezembro de 2023 recolhidos a cada 15 minutos no contador da E-Redes (empresa distribuidora de eletricidade em Portugal Continental). Os dados em questão foram disponibilizados pelos professores orientadores, em ficheiros CSV, contendo os registos de data e hora e o valor da energia consumida, medido em quilowatt-hora (kWh). Apesar da frequência regular de 15 minutos, nem todos os dias do período analisado contêm registos completos, isto é, verificou-se a existência de dias incompletos no início e no final do período, os quais foram removidos durante a fase da preparação dos dados, conforme descrito na secção seguinte.

4.2. Preparação do *Dataset*

A análise teve início na importação da base de dados mencionada anteriormente, contendo o registo do consumo energético em intervalos de 15 minutos. Procedeu-se à normalização do nome das colunas para facilitar o manuseamento, seguido da conversão da coluna DataHora para o tipo *datetime*. Esta coluna é então definida como índice do *dataframe*, permitindo uma manipulação temporal dos dados com maior eficiência.

De modo a simplificar as análises e alinhar os dados com escalas temporais mais convencionais, foi realizada uma agregação horária, somando os valores de cada hora. Esta transformação converte o *dataset* original em base horária, mais apropriada para análises de padrões ao longo do tempo.

4.2.1. Filtragem de Dias Completos

Para garantir a integridade das análises subsequentes, foram removidos os dias incompletos, ou seja, dias que não apresentavam registos para as 24 horas. Esta filtragem foi realizada através da contagem de entradas por dia e da seleção exclusiva dos dias com 24 registos

completos. Desta forma, eliminaram-se eventuais distorções causadas por dias truncados no início ou no final da amostragem.

4.2.2. Criação do *Dataset* em Base Diária

Após a limpeza do *dataset* original em base horária, procedeu-se à criação de um novo *dataset* em base diária. Esta transformação foi realizada através da agregação dos valores horários, utilizando a soma do consumo energético ao longo de cada período de 24 horas consecutivas. O *dataset* resultante foi posteriormente utilizado em etapas subsequentes do estudo, nomeadamente no treino e avaliação dos modelos de previsão.

4.3. Análise Exploratória dos Dados

Com os dados limpos e estruturados, foram geradas várias visualizações para compreender o comportamento do consumo energético ao longo do tempo.

4.3.1. Evolução Temporal do Consumo

A série temporal horária, apresentada na Figura 12 permite observar o comportamento do consumo energético do *campus* entre 2015 e 2023. O gráfico evidencia a presença de padrões sazonais ao longo dos anos, com períodos de maior consumo que poderão estar associados a variações sazonais na utilização das instalações e, possivelmente, à maior utilização de sistemas de climatização.

Observa-se ainda uma redução significativa do consumo a partir de 2020, consistente com os impactos da pandemia de COVID-19 e com a consequente diminuição da atividade presencial no *campus*. Nos anos subsequentes, observa-se uma recuperação gradual do consumo, embora os valores se mantenham em níveis inferiores aos registados antes da pandemia. Esta redução poderá estar associada a alterações nos padrões de utilização das instalações, nomeadamente a uma maior adoção do trabalho remoto e a mudanças de funcionamento do *campus*, que se terão mantido mesmo após o levantamento das medidas de controlo da pandemia.

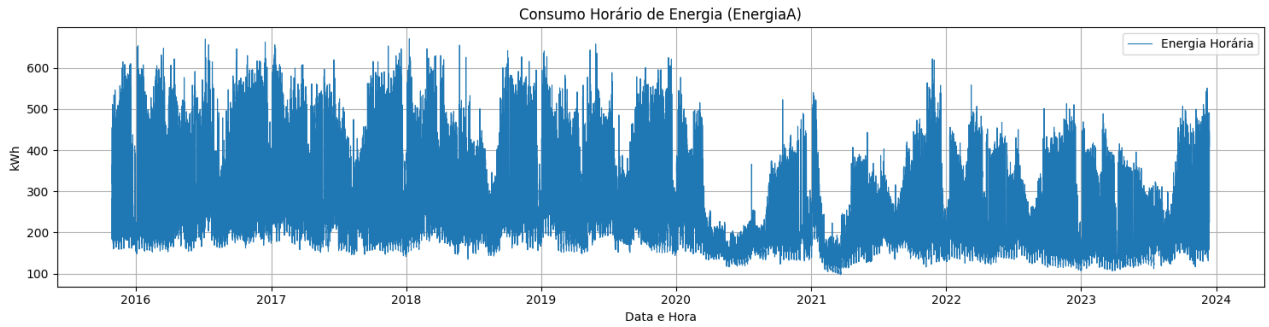

Figura 12: Evolução do Consumo de Energia (2015 a 2023)

4.3.2. Perfil Médio Diário de Consumo

O perfil médio diário, apresentado na Figura 13, apresenta a média do consumo por hora ao longo de um dia típico. Observa-se um aumento acentuado entre as 07h e as 11h, com um pico máximo próximo do meio-dia. O consumo mantém-se elevado durante a tarde e começa a decrescer a partir das 18h. Durante a madrugada, os valores mantêm-se estáveis e reduzidos, evidenciando um perfil de consumo diurno, característico de edifícios administrativos, escolares ou de serviços.

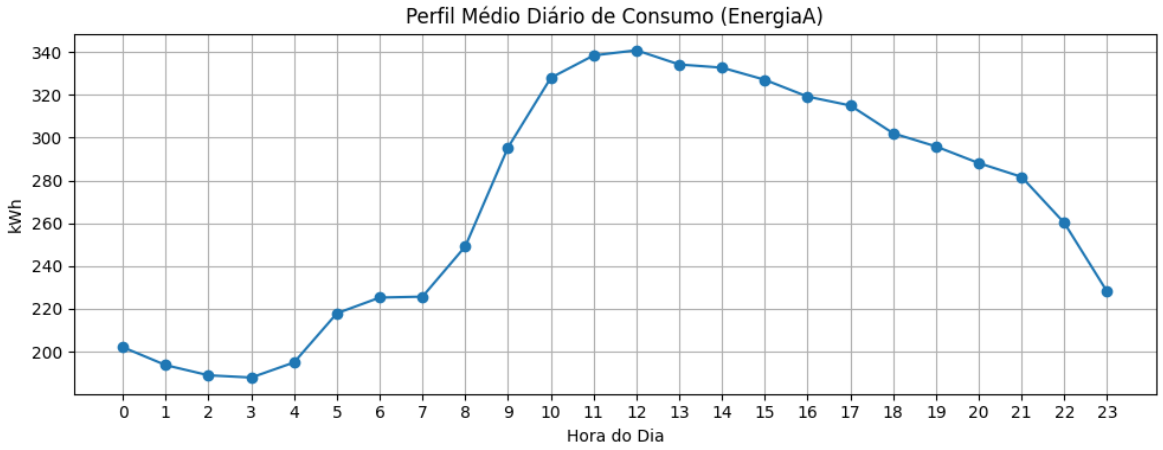

Figura 13: Perfil Médio Diário de Consumo

4.3.3. Distribuição do Consumo

Na Figura 14, verifica-se a distribuição do consumo energético ao longo dos diferentes meses do ano, permitindo analisar a variabilidade e a dispersão dos consumos em cada período. Constata-se que os valores médios de consumo apresentam variações ao longo do ano, sendo, em geral, mais elevados nas estações do outono e inverno, nomeadamente em janeiro, outubro e novembro. Por outro lado, observa-se uma redução significativa do

consumo durante o período de verão, com destaque para o mês de agosto, que apresenta o valor médio mais reduzido. Este comportamento é consistente com a variação sazonal das condições climáticas ao longo do ano, uma vez que, durante os meses de inverno, a redução das horas de luz natural poderá conduzir a uma maior utilização de iluminação artificial e, em alguns casos, à utilização de equipamentos adicionais, como aquecedores individuais, contribuindo para um aumento do consumo energético. Por outro lado, durante o período de verão, e em particular no mês de agosto, a redução da atividade associada à ausência de aulas ou funcionamento reduzido permite justificar a diminuição dos níveis de consumo observados. Adicionalmente, observa-se a existência de alguma variabilidade nos valores de consumo entre os diferentes meses, bem como a presença de valores extremos em determinados períodos, indicando a ocorrência ocasional de picos de consumo ao longo do ano. Em termos de amplitude total da variação dos dados, observa-se que, na maioria dos meses, os valores mínimos de consumo se situam em torno dos 100 kWh, enquanto os valores máximos podem atingir aproximadamente 650 a 700 kWh. No entanto, no mês de agosto, o valor máximo situa-se próximo dos 500 kWh, consideravelmente abaixo dos restantes meses, refletindo a redução generalizada do consumo durante este período.

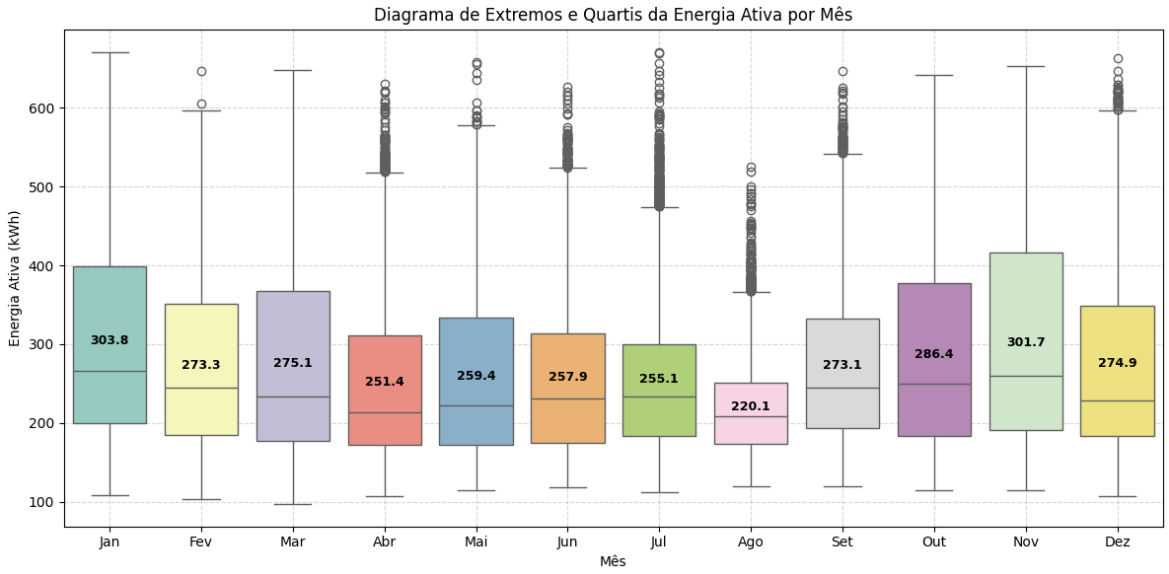

Figura 14: Diagrama de Extremos e Quartis do consumo energético por mês. Os valores apresentados a negrito no interior das caixas correspondem ao valor médio mensal do consumo energético.

Observando a Figura 15, verifica-se a distribuição do consumo energético ao longo dos diferentes dias da semana, permitindo analisar não apenas os valores médios, mas também a variabilidade e a dispersão dos consumos. Constata-se que os dias úteis, entre segunda-feira e sexta-feira, apresentam valores médios de consumo relativamente semelhantes, situando-

-se em torno dos 300 kWh, refletindo um padrão consistente de utilização associado aos períodos de maior atividade. Em contraste, durante o fim de semana observa-se uma redução significativa dos valores de consumo, com valores médios mais baixos ao sábado, e, sobretudo, ao domingo, evidenciando uma diminuição da atividade associada aos consumos energéticos. Adicionalmente, verifica-se a presença de alguns valores extremos ao longo dos diferentes dias da semana, indicando a ocorrência ocasional de picos de consumo que poderão estar associados a situações específicas de funcionamento dos sistemas.

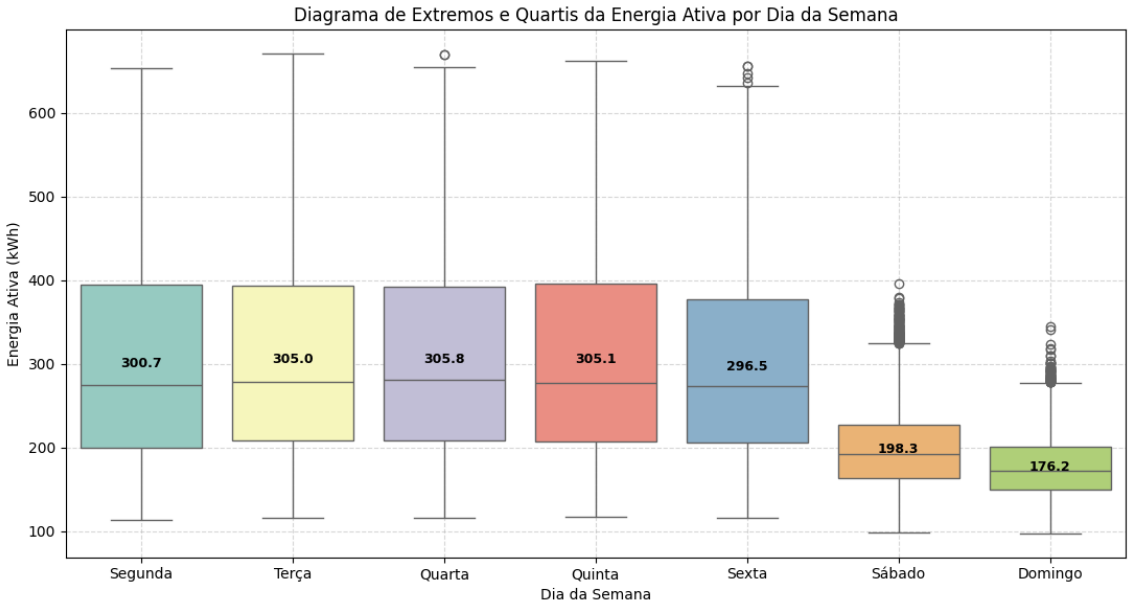

Figura 15: Diagrama de Extremos e Quartis por Semana

5. Desenvolvimento de Modelos

O presente capítulo descreve o processo de desenvolvimento, implementação e avaliação dos modelos de estimação do consumo energético. São apresentados os procedimentos adotados para a construção dos seis modelos de ML definidos no Capítulo 2, bem como a definição dos seus hiperparâmetros.

5.1. Construção de Conjuntos de Dados

Nesta etapa, os dados foram preparados para o desenvolvimento e avaliação dos modelos preditivos, assegurando a sua correta organização temporal e adequação ao processo de modelação. O conjunto de dados foi dividido em treino, validação e teste com base em períodos consecutivos, garantindo que não era utilizada informação de períodos futuros no treino dos modelos. A Tabela 1 apresenta a divisão temporal utilizada neste estudo.

Tabela 1: Divisão dos Dados

	Data de Início	Data de Fim	Nº de Observações
Treino	28/10/2015	13/12/2021	53.736
Validação	14/12/2021	13/12/2022	8.760
Teste	14/12/2022	13/12/2023	8.760

Após a divisão dos dados, foi aplicada uma normalização utilizando o método *Min-Max Scaling*, com o objetivo de garantir que a variável de consumo energético apresentasse valores numa escala adequada, facilitando o processo de aprendizagem.

De forma a preparar os dados para os modelos baseados em séries temporais, foram criadas sequências de entrada utilizando uma janela temporal deslizante de 24 horas. Cada sequência foi composta pelos valores observados nas 24 horas anteriores, sendo utilizada para prever o valor imediatamente seguinte. Este procedimento permitiu transformar a série temporal original num conjunto de exemplos supervisionados adequado ao treino dos modelos.

A mesma metodologia de preparação dos dados foi aplicada a todos os modelos considerados neste estudo, garantindo a consistência experimental e a comparabilidade dos resultados obtidos. Para cada modelo, foram testadas diferentes combinações de parâmetros, com o objetivo de identificar as configurações que proporcionassem o melhor desempenho no conjunto de validação.

5.2. Otimização de Hiperparâmetros

De forma a melhorar o desempenho dos modelos preditivos desenvolvidos neste estudo, foi realizado um processo de otimização de hiperparâmetros utilizando a biblioteca *Optuna*, uma ferramenta amplamente utilizada para a seleção automática de parâmetros em modelos de aprendizagem computacional. O objetivo deste processo consistiu em identificar as combinações de parâmetros que proporcionassem o melhor desempenho preditivo para cada modelo considerado. Para tal, foram definidas diferentes configurações possíveis para os principais hiperparâmetros de cada arquitetura, incluindo, entre outros, o número de unidades nas camadas, a taxa de aprendizagem (*learning rate*), o tamanho do *batch* e a taxa de *dropout*. A otimização foi conduzida através da execução de múltiplos ensaios (*trials*).

Durante o treino dos modelos, foi utilizado o MAPE como critério de avaliação, sendo o objetivo minimizar o erro obtido no conjunto de validação. Adicionalmente, foi aplicado o mecanismo de *Early Stopping*, permitindo interromper automaticamente o treino quando não se verificasse melhoria no desempenho do modelo após um determinado número de épocas, contribuindo para reduzir o risco de *overfitting* e otimizar o tempo de treino. Importa referir que o número máximo de épocas de treino foi previamente definido e mantido ao longo do processo, não sendo incluído como hiperparâmetro no procedimento de otimização.

Os intervalos e valores testados para cada hiperparâmetro encontram-se descritos em detalhe no Anexo A.

5.3. Estimação do Consumo Energético

Após a identificação dos melhores hiperparâmetros, procedeu-se à implementação, treino e avaliação dos modelos considerados, com o objetivo de analisar o seu comportamento preditivo. Foi seguido um procedimento comum para todas as arquiteturas, garantindo a consistência metodológica e a comparabilidade dos resultados obtidos.

Importa clarificar a distinção terminológica adotada ao longo deste trabalho entre os conceitos de previsão e estimação. O termo **previsão** é utilizado para descrever a obtenção de valores futuros de consumo energético, enquanto o termo **estimação** é utilizado para descrever o cálculo do consumo energético diário obtido por agregação das previsões horárias ao longo de um dia completo. Nesta abordagem, a estimativa diária apenas pode ser determinada após a conclusão do período temporal considerado, isto é, no final do dia. Assim, a metodologia baseada na agregação das previsões horárias não constitui um

mecanismo de previsão direta de valores futuros em base diária, mas sim um procedimento de estimação do consumo diário.

5.3.1. Dados Horários

Inicialmente, foi realizada a importação da base de dados horária, correspondente à variável de consumo energético em estudo. Os dados foram normalizados e transformados em sequências temporais através de uma janela deslizante de 24 horas. Com base nos melhores hiperparâmetros identificados, procedeu-se à construção e treino de cada modelo utilizando o conjunto de treino, monitorizando o desempenho no conjunto de validação. O treino foi realizado com recurso à função de perda MSE, sendo também acompanhada a métrica MAPE. Adicionalmente, foi utilizado o mecanismo de *Early Stopping*, com o objetivo de interromper o treino quando não se verificassem melhorias no desempenho do modelo no conjunto de validação.

Após concluído o treino, cada modelo foi utilizado para gerar previsões no conjunto de teste. Uma vez que os dados se encontravam normalizados, as previsões obtidas foram convertidas para a escala original, de modo a permitir uma interpretação direta dos resultados e o cálculo das métricas de erro em unidades reais de consumo energético. Com base nas previsões horárias e nos respetivos valores observados, foram então calculadas as métricas de avaliação de desempenho, nomeadamente o MSE, o RMSE, MAPE, MAE e R^2.

Numa etapa posterior, as previsões horárias foram agregadas para uma base diária, através da soma das 24 observações horárias correspondentes a cada dia, obtendo estimativas do consumo energético total diário, sobre as quais foram calculadas as métricas de avaliação de desempenho em base diária.

De forma a avaliar a estabilidade e a robustez dos resultados, foi implementado um procedimento de execução repetida do processo de treino e avaliação. Para cada arquitetura, foram realizados 50 ciclos independentes, mantendo a configuração de hiperparâmetros previamente definida. Em cada ciclo, foi utilizado um valor distinto de *seed*, garantindo variações na inicialização dos pesos e do processo de aprendizagem. O modelo foi então recriado e treinado seguindo o mesmo procedimento descrito anteriormente, sendo as previsões horárias geradas, convertidas para a escala original e agregadas para a base diária, possibilitando o cálculo das métricas de avaliação em ambas as granularidades.

5.3.2. Dados Diários

Foi também realizada uma abordagem de previsão direta utilizando os dados em base diária. Para tal, os dados foram normalizados e transformados em sequências temporais através de uma janela deslizante de sete dias. Foram seguidos os mesmos procedimentos de treino e avaliação descritos para os dados horários, aplicando-se a mesma arquitetura e configurações de hiperparâmetros. As previsões obtidas foram convertidas para a escala original dos dados, possibilitando o cálculo das métricas de avaliação do desempenho.

6. Análise dos Resultados

Neste capítulo são apresentados e analisados os resultados obtidos pela aplicação dos seis modelos de estimação considerados neste trabalho. Os resultados apresentados resultam da execução dos modelos ao longo dos 50 ciclos independentes de treino, realizados com diferentes sementes aleatórias (*seeds*), permitindo uma avaliação mais consistente do seu desempenho. Em cada ciclo, as métricas de avaliação foram calculadas na amostra de teste, sendo posteriormente apresentados a média e o desvio padrão obtidos ao longo dos 50 ciclos. Os modelos foram avaliados na granularidade horária e na estimativa do consumo diário por agregação das previsões horárias. Para contextualização, são também brevemente mencionados os resultados obtidos com a abordagem de previsão direta em dados diários, não sendo estas abordagens diretamente comparáveis entre si, por utilizarem granularidades e informação de entrada distintas.

6.1. Desempenho dos Modelos de Previsão Horária

A Tabela 2 apresenta os resultados médios e os respetivos desvios padrão obtidos pelos seis modelos na granularidade horária. As métricas foram calculadas na amostra de teste em cada um dos 50 ciclos independentes de treino.

Tabela 2: Média e desvio padrão das métricas de avaliação calculadas na amostra de teste, em base horária, ao longo dos 50 ciclos independentes de treino

Modelo	RMSE (kWh) Média ± DP	MAE (kWh) Média ± DP	MAPE (%) Média ± DP (%)	R^2 Média ± DP
MLP	17,21 ± 1,32	12,46 ± 0,88	5,70 ± 0,36	0,9627 ± 0,0060
LSTM	15,84 ± 0,43	11,47 ± 0,38	5,25 ± 0,22	0,9685 ± 0,0017
GRU	16,11 ± 0,90	11,65 ± 0,72	5,32 ± 0,36	0,9674 ± 0,0039
N-BEATS	19,90 ± 0,65	13,04 ± 0,45	5,83 ± 0,21	0,9506 ± 0,0032
XGBoost	17,24 ± 0,04	12,32 ± 0,03	5,69 ± 0,01	0,9628 ± 0,0002
N-HiTS	26,64 ± 2,27	19,05 ± 1,91	9,36 ± 0,84	0,9109 ± 0,0154

O modelo LSTM destacou-se como o mais preciso nesta granularidade, apresentando o menor valor médio de MAPE (5,25% ± 0,22%), bem como o maior valor médio de R^2 (0,9685 ± 0,0017). O modelo GRU apresentou um modelo semelhante, com valores de erro

e de R^2 próximos dos obtidos pelo LSTM, sugerindo que ambas as arquiteturas recorrentes são adequadas para a modelação de séries temporais de consumo energético. Os modelos MLP, XGBoost e N-BEATS apresentaram desempenhos idênticos entre si, com valores de MAPE situados entre 5,69% e 5,83%. Destaca-se, no entanto, o modelo XGBoost pela sua elevada estabilidade, evidenciada por um desvio padrão bastante reduzido (0,01%), refletindo a sua consistência entre execuções. Em contraste, o modelo N-HiTS registou o desempenho menos favorável em todas as métricas analisadas, apresentando o maior valor médio de MAPE (9,36% ± 0,84%) e o menor valor médio de R^2 (0,9109 ± 0,0154), bem como a maior variabilidade entre ciclos.

6.2. Desempenho dos Modelos de Agregação Diária

A Tabela 3 apresenta os resultados médios e os respetivos desvios padrão obtidos após a agregação das previsões horárias para base diária, através da soma das 24 previsões horárias correspondentes a cada dia. As métricas foram calculadas na amostra de teste em cada um dos 50 ciclos independentes de treino.

Tabela 3: Média e desvio padrão das métricas de avaliação calculadas na amostra de teste, para a estimativa diária por agregação horária, ao longo dos 50 ciclos de treino

Modelo	RMSE (kWh) Média ± DP	MAE (kWh) Média ± DP	MAPE (%) Média ± DP (%)	R^2 Média ± DP
MLP	109,30 ± 22,62	86,87 ± 19,21	1,70 ± 0,39	0,9946 ± 0,0023
LSTM	89,34 ± 13,10	69,92 ± 11,76	1,37 ± 0,28	0,9964 ± 0,0011
GRU	103,18 ± 23,76	81,21 ± 19,67	1,57 ± 0,38	0,9951 ± 0,0025
N-BEATS	90,02 ± 11,35	67,87 ± 10,54	1,28 ± 0,18	0,9964 ± 0,0009
XGBoost	90,42 ± 0,67	69,76 ± 0,61	1,49 ± 0,01	0,9964 ± 0,0001
N-HiTS	362,69 ± 40,12	258,22 ± 40,68	5,69 ± 0,69	0,9421 ± 0,0127

Na estimativa diária por agregação, o modelo N-BEATS obteve o melhor valor médio de MAPE (1,28% ± 0,18), bem como valores reduzidos de RMSE e MAE, evidenciando um elevado desempenho preditivo nesta abordagem. O modelo LSTM apresentou um desempenho muito semelhante, com um MAPE médio de 1,37%, seguido do XGBoost (1,49%) e do GRU (1,57%). O modelo XGBoost volta a destacar-se pela sua elevada estabilidade, mesmo após a agregação dos dados, apresentando um desvio padrão muito

reduzido (0,01%). Em contraste, e à semelhança do observado na previsão em base horária, o modelo N-HiTS apresentou o desempenho menos favorável, com valores de erro substancialmente superiores aos restantes modelos, nomeadamente um MAPE de 5,69%, indicando uma menor capacidade de generalização.

Para complementar a análise das médias, a Tabela 4 apresenta os melhores resultados individuais obtidos por cada modelo ao longo dos 50 ciclos independentes, identificados com base no menor MAPE da estimação diária. Esta análise permite avaliar o potencial máximo de cada arquitetura e verificar se as diferenças entre os modelos se mantêm quando se considera o seu melhor desempenho possível.

Tabela 4: Melhores resultados individuais de cada modelo ao longo dos 50 ciclos na estimação diária

Modelo	RMSE (kWh)	MAE (kWh)	MAPE (%)	R^2
MLP	76,65	53,71	1,00	0,9974
LSTM	69,39	51,88	0,96	0,9979
GRU	75,05	53,58	0,97	0,9975
N-BEATS	73,01	52,32	0,98	0,9977
XGBoost	88,90	68,77	1,46	0,9966
N-HiTS	256,51	170,57	3,95	0,9714

O LSTM obteve o melhor resultado individual na estimativa diária, com um MAPE de 0,96% e R^2 de 0,9979, face a um valor médio de 1,37% ± 0,29% ao longo das diferentes execuções. O modelo GRU e o N-BEATS apresentaram desempenhos semelhantes, com valores de MAPE de 0,97% e 0,98%, respetivamente, face a médias globais de 1,57% e 1,28%. O MLP registou igualmente um resultado individual relevante, com MAPE de 1,00%, comparativamente a uma média de 1,70%. Por sua vez, o modelo XGBoost apresentou um comportamento distinto: o seu melhor resultado individual (1,46%) relevou-se muito próximo do valor médio obtido (1,49% ± 0,01%), refletindo a elevada estabilidade deste algoritmo. O modelo N-HiTS manteve-se como a abordagem com desempenho menos favorável, mesmo no seu melhor ciclo (3,95%).

Para ilustrar visualmente este resultado, apresenta-se na Figura 16 a evolução dos valores reais e estimados do consumo diário ao longo do período de teste, considerando o modelo

LSTM no seu melhor ciclo. Verifica-se uma elevada proximidade entre os valores estimados e os valores reais ao logo do período de teste, evidenciando a capacidade do modelo em reproduzir o comportamento global do consumo energético diário.

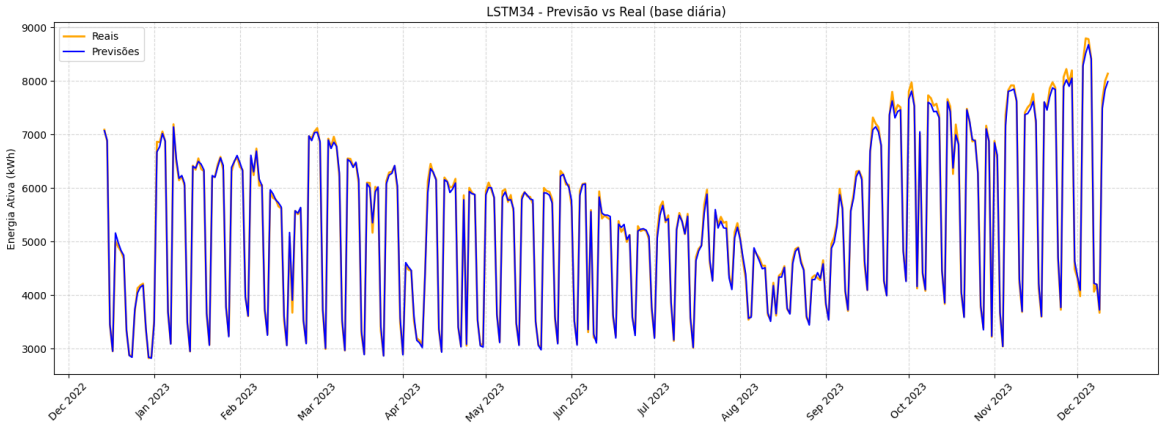

Figura 16: Consumo diário real e estimado pelo modelo LSTM no seu melhor ciclo individual, obtido por agregação das previsões horárias.

6.3. Discussão

Os resultados obtidos ao longo dos 50 ciclos independentes de treino demonstram que a metodologia de estimação proposta, baseada na previsão horária do consumo energético com posterior agregação para a base diária, permite obter valores de erro extremamente baixos e consistentes. Os cinco melhores modelos registaram MAPE médio diário entre 1,28% (N-BEATS) e 1,70% (MLP), com o melhor resultado individual a atingir 0,96% no caso do LSTM. Para contextualizar estes resultados, importa referir que os modelos treinados diretamente com dados diários registaram valores de MAPE entre 8,16% (N-HiTS) e 11,21% (XGBoost), valores substancialmente superiores aos obtidos pela estimativa por agregação horária. A Figura 17 apresenta a evolução dos valores reais e previstos do consumo diário ao longo do período de teste, ilustrando o desempenho do modelo N-HiTS treinado diretamente em base diária. Em comparação com a Figura 16, observa-se uma menor proximidade entre os valores previstos e os valores reais, refletindo o desempenho inferior do modelo treinado diretamente em base diária.

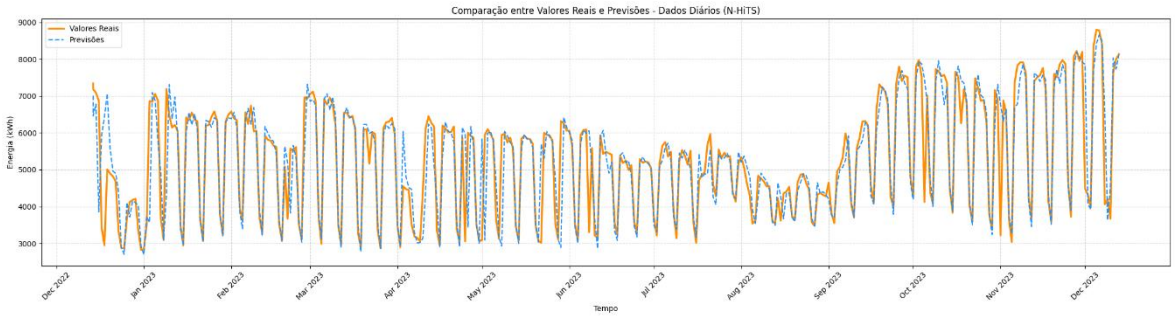

Figura 17: Consumo diário real e previsto pelo modelo N-HiTS treinado diretamente em base diária.

Esta diferença, ainda que não diretamente comparável, dado que as abordagens operam sobre granularidades distintas e que uma se destina à estimação do consumo num dia que já terminou e outra à previsão do consumo do dia seguinte, evidencia a relevância da granularidade horária na qualidade da estimação diária. A redução do erro observada nesta metodologia pode ser explicada pelo efeito de compensação dos erros individuais ao longo do dia, uma vez que as previsões horárias sucessivas, quando agregadas, tendem a reduzir o impacto de desvios pontuais, resultando numa estimativa diária globalmente mais precisa. Para ilustrar este efeito de forma quantitativa, calculou-se, para um dia representativo do conjunto de teste, o somatório do módulo dos erros horários e o somatório dos erros horários sem módulo, cujos valores se encontram ilustrados na Figura 18. O somatório em valor absoluto atinge 93,96 kWh, enquanto o somatório sem módulo é de apenas 16,08 kWh, evidenciando que os erros positivos (a vermelho) e negativos (a verde) tendem a compensar-se mutuamente ao longo das 24 horas.

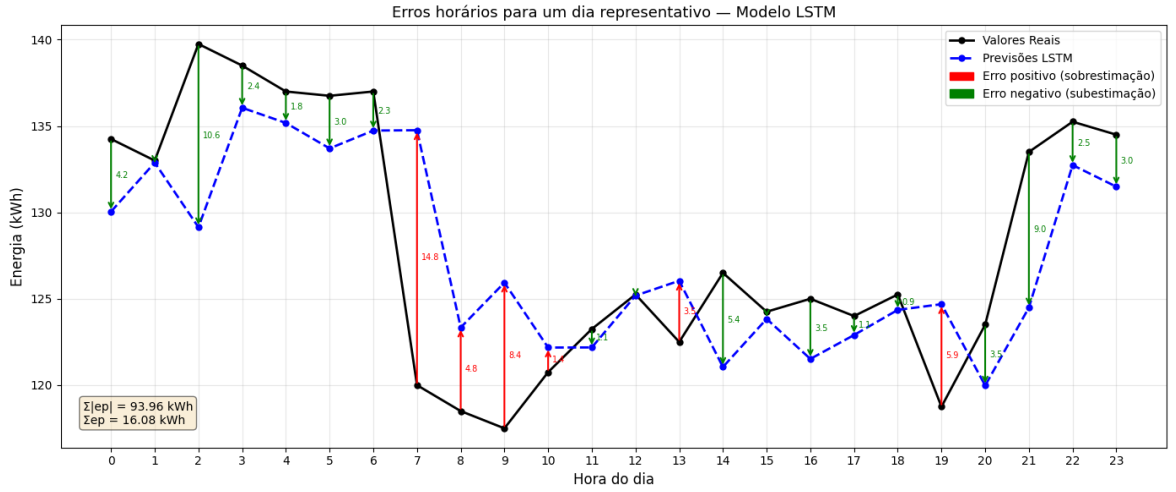

Figura 18: Erros horários para um dia representativo do conjunto de teste, obtidos com o modelo LSTM

Uma observação adicional prende-se com a convergência dos valores de MAPE entre modelos de arquiteturas distintas. Tanto na previsão horária como na estimação diária, os cinco melhores modelos registaram resultados de ordem de grandeza muito semelhante. A título de hipótese, esta proximidade poderá indicar que os modelos se aproximam do limite de desempenho alcançável com os dados e a metodologia disponíveis, sugerindo que eventuais ganhos adicionais poderão requerer a incorporação de variáveis exógenas ou a adoção de abordagens metodológicas distintas.

Os valores de erro baixos obtidos, constituem uma referência de consumo esperado suficientemente precisa para suportar a identificação de desvios significativos no consumo energético em edifícios institucionais. A deteção de anomalias assenta precisamente na comparação entre o consumo diário estimado e o consumo efetivamente observado, sendo que desvios que ultrapassem um determinado limiar podem ser sinalizados como potencialmente anómalos. A definição desse limiar depende do contexto operacional e do conhecimento do domínio, estando fora do âmbito específico deste trabalho.

7. Conclusão

O presente trabalho teve como objetivo o desenvolvimento e avaliação de modelos de estimação do consumo energético com elevado nível de precisão, com vista à sua utilização como referência de consumo esperado no contexto da deteção de comportamento anómalos em edifícios institucionais. Dando continuidade ao trabalho desenvolvido por J. Sá [6], que havia identificado o potencial desta abordagem sem a ter explorado de forma sistemática, este estudo focou-se na obtenção de modelos com erro de estimação muito reduzido, condição necessária para que os desvios entre os valores estimados e os valores observados possam ser interpretados como indicadores fiáveis de comportamentos anómalos.

Foram desenvolvidos e avaliados seis modelos distintos, treinados ao longo de 50 ciclos independentes, de forma a avaliar o seu desempenho preditivo e a robustez dos resultados obtidos. Na granularidade horária, o LSTM destacou-se como o modelo mais preciso, registando um MAPE médio de 5,25%, sendo o GRU o modelo com desempenho mais próximo, com 5,32%. No que respeita à estimativa diária por agregação, o N-BEATS evidenciou o melhor desempenho médio, com um MAPE de 1,28%, enquanto o LSTM apresentou 1,37%. A redução do erro observada nesta metodologia resulta do efeito de compensação dos erros individuais ao longo do dia, uma vez que ao agregar as previsões horárias sucessivas, os desvios positivos e negativos tendem a anular-se, resultando numa estimativa diária globalmente mais precisa. Esta abordagem permite, assim, transformar a informação acumulada ao longo do dia numa referência de consumo com elevado grau de precisão.

Em termos de robustez, o XGBoost revelou-se o modelo mais estável, com desvios padrão de apenas 0,01% no MAPE, tanto na previsão dos dados horários, como na estimação dos dados diários. O LSTM e o N-BEATS afirmaram-se como as arquiteturas mais equilibradas, combinando bom desempenho médio com variabilidade controlada, tendo o LSTM registado um MAPE diário médio de 1,37% ± 0,28% e o N-BEATS de 1,28% ± 0,18%. Por sua vez, o N-HiTS apresentou consistentemente o pior desempenho e a maior variabilidade do conjunto, com um MAPE horário de 9,36% ± 0,84% e um MAPE diário de 5,69% ± 0,70%, valores substancialmente superiores aos restantes modelos.

Os resultados obtidos superam os alcançados pelo trabalho de J. Sá [6], cujo melhor modelo RNN registou um MAPE de 1,82% na estimação diária por agregação horária e um MAPE médio de 2,82% ao longo de 30 ciclos.

7.1. Limitações

O desenvolvimento deste trabalho foi condicionado por algumas limitações que importa referir. A principal limitação prende-se com os recursos computacionais disponíveis, uma vez que todo o processamento foi realizado num computador pessoal, sem acesso a unidades de processamento gráfico (GPU) ou a recursos computacionais adicionais. Esta condicionante teve impacto, em primeiro lugar, no processo de otimização de hiperparâmetros. Devido às limitações de memória e capacidade de processamento, foi necessário realizar vários ensaios de otimização de forma separada, uma vez que a execução simultânea de um número elevado de combinações de parâmetros se revelou inviável nas condições disponíveis. Em segundo lugar, estas limitações tiveram impacto no tempo de execução dos ciclos independentes de treino, tornando exigente a realização de um número superior de ciclos por modelo. Embora tenham sido realizados 50 ciclos independentes, um maior número de execuções permitiria aprofundar a análise de robustez dos resultados obtidos. Por fim, verificaram-se diferenças no tempo de execução entre modelos, sendo o N-HiTS a arquitetura mais exigente do ponto de vista computacional, o que condicionou a exploração de um maior número de configurações de hiperparâmetros para este modelo.

7.2. Trabalho futuro

Os resultados obtidos abrem várias perspetivas de desenvolvimento futuro. Em primeiro lugar, a implementação efetiva de um sistema de deteção de anomalias com base nos desvios gerados pelos modelos desenvolvidos constitui o passo natural de continuação deste trabalho. A definição de limiares de alerta, eventualmente ajustados por período do dia, dia de semana ou estação do ano, permitiria transformar os modelos desenvolvidos numa ferramenta operacional de apoio à gestão energética dos edifícios.

Em segundo lugar, a incorporação de variáveis exógenas, como a temperatura, a ocupação do edifício ou o calendário de atividades, poderá contribuir para reduzir ainda mais o erro de previsão, particularmente em períodos de maior variabilidade sazonal. O trabalho do J. Sá [6] demonstrou que a inclusão de variáveis de ocupação melhora significativamente o

desempenho dos modelos, abordagem que não foi explorada no presente trabalho e que poderá beneficiar as arquiteturas aqui avaliadas.

Por último, a validação da metodologia proposta em edifícios com perfis de consumo distintos e a sua eventual integração com sistemas de gestão energética em tempo real constituem etapas importantes para a aplicação prática dos modelos desenvolvidos.

Bibliografia

[1] Observatório da Energia, DGEG – Direção Geral de Energia e Geologia, Direção de Serviços de Planeamento Energético e Estatística, e ADENE – Agência para a Energia, Direção de Formação, Informação e Educação, «Energia em Números – Edição 2024», Lisboa, jun. 2024. [Online]. Disponível em: https://www.dgeg.gov.pt/pt/destaques/energia-em-numeros-edicao-2024/

[2] Q. Zeng, F. Peng, e X. Han, «Towards sustainable architecture: Enhancing green building energy consumption prediction with integrated variational autoencoders and self-attentive gated recurrent units from multifaceted datasets», *PLOS ONE*, vol. 20, n.º 4, p. e0317514, abr. 2025, doi: https://doi.org/10.1371/journal.pone.0317514.

[3] Y. Yang, Q. Duan, e F. Samadi, «A systematic review of building energy performance forecasting approaches», *Renewable and Sustainable Energy Reviews*, vol. 223, p. 116061, nov. 2025, doi: 10.1016/j.rser.2025.116061.

[4] Z. Mao, B. Zhou, J. Huang, D. Liu, e Q. Yang, «Research on Anomaly Detection Model for Power Consumption Data Based on Time-Series Reconstruction», *Energies*, vol. 17, n.º 19, p. 4810, set. 2024, doi: 10.3390/en17194810.

[5] D. Stjelja, V. Kuzmanovski, R. Kosonen, e J. Jokisalo, «Building consumption anomaly detection: A comparative study of two probabilistic approaches», *Energy and Buildings*, vol. 313, p. 114249, jun. 2024, doi: 10.1016/j.enbuild.2024.114249.

[6] J. de Sá, «Aplicação de Técnicas de Ciência de Dados na Previsão de Consumos Energéticos», Master's thesis, Instituto Politécnico de Leiria, Leiria, Portugal, Setembro 2023. [Online]. Disponível em: http://hdl.handle.net/10400.8/9525.

[7] A. M. Shimaoka, R. C. Ferreira, e A. Goldman, «The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks», *Reviews*, vol. 4, n.º 1, pp. 28–43, out. 2024, doi: 10.5753/reviews.2024.3757.

[8] J. X. Leon-Medina *et al.*, «Forecasting Energy Demand in Quicklime Manufacturing: A Data-Driven Approach», *Sensors*, vol. 25, n.º 24, p. 7632, dez. 2025, doi: 10.3390/s25247632.

[9] F. Zhang, C. Deb, S. E. Lee, J. Yang, e K. W. Shah, «Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique», *Energy and Buildings*, vol. 126, pp. 94–103, ago. 2016, doi: 10.1016/j.enbuild.2016.05.028.

[10] Q. Qiao, A. Yunusa-Kaltungo, e R. E. Edwards, «Feature selection strategy for machine learning methods in building energy consumption prediction», *Energy Reports*, vol. 8, pp. 13621–13654, nov. 2022, doi: 10.1016/j.egyr.2022.10.125.

[11] S. F. Ardabili, L. Abdilalizadeh, C. Mako, B. Torok, e A. Mosavi, «Systematic review of deep learning and machine learning for building energy», 23 de fevereiro de 2022, *Open Science Framework*. doi: 10.31219/osf.io/fxtmz.

[12] M. F. Manzoor, «Energy Load Forecasting with Machine Learning: Models, Metrics, and Future Directions», *PJAI*, ago. 2025, doi: 10.70389/PJAI.100018.

[13] C. M. Viana, S. Oliveira, e J. Rocha, «Introductory Chapter: Time Series Analysis», em *Time Series Analysis - Recent Advances, New Perspectives and Applications*, J. Rocha, C. M. Viana, e S. Oliveira, Eds., IntechOpen, 2024. doi: 10.5772/intechopen.1004609.

[14] K. Guedes, «Análise de séries temporais: conceitos e aplicações», Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação), Universidade Federal do Maranhão, São Luís, MA, Brasil, 2025.

[15] F. S. Campos, «Previsão de séries temporais de pressão em redes de distribuição de água aplicando modelos generalizados autorregressivos de médias móveis (GARMA)», Mestrado em Hidráulica e Saneamento, Universidade de São Paulo, São Carlos, 2023. doi: 10.11606/D.18.2023.tde-08012024-111307.

[16] K. Kawa, R. Mularczyk, W. Bauer, K. Grobler-Dębska, e E. Kucharska, «Prediction of Energy Consumption on Example of Heterogenic Commercial Buildings», *Energies*, vol. 17, n.º 13, p. 3220, jun. 2024, doi: 10.3390/en17133220.

[17] R. J. Hyndman e G. Athanasopoulos, *Forecasting: Principles and Practice*, 3.ª ed. Monash University, 2021. [Online]. Disponível em: https://otexts.com/fpp3/

[18] N. Maragkos e I. Refanidis, «A Comparative Evaluation of Time-Series Forecasting Models for Energy Datasets», *Computers*, vol. 14, n.º 7, p. 246, jun. 2025, doi: 10.3390/computers14070246.

[19] M. Karatzoglidi, P. Kerasiotis, e V. Kantere, «Automated energy consumption forecasting with EnForce», *Proc. VLDB Endow.*, vol. 14, n.º 12, pp. 2771–2774, jul. 2021, doi: 10.14778/3476311.3476341.

[20] S. Sathyanarayana e T. Mohanasundaram, «Stationarity and Unit Roots in Time Series: Theoretical Insights and Practical Considerations», *IRAJMSS*, vol. 21, n.º 2, p. 46, ago. 2025, doi: 10.21013/jmss.v21.n2.p1.

[21] S. Knox e M. Smal, «The seasonal adjustment methodology applied to economic statistics by the South African Reserve Bank», 2023.

[22] G. Dudek, «STD: A Seasonal-Trend-Dispersion Decomposition of Time Series», 21 de abril de 2022, *arXiv*: arXiv:2204.10398. doi: 10.48550/arXiv.2204.10398.

[23] K. Kyo, H. Noda, e F. Fang, «An integrated approach for decomposing time series data into trend, cycle and seasonal components», *Mathematical and Computer Modelling of*

Dynamical Systems, vol. 30, n.º 1, pp. 792–813, dez. 2024, doi: 10.1080/13873954.2024.2416631.

[24] N. Maleki, O. Lundström, A. Musaddiq, J. Jeansson, T. Olsson, e F. Ahlgren, «Future energy insights: Time-series and deep learning models for city load forecasting», *Applied Energy*, vol. 374, p. 124067, nov. 2024, doi: 10.1016/j.apenergy.2024.124067.

[25] J. Kim, H. Kim, H. Kim, D. Lee, e S. Yoon, «A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges», *Artif Intell Rev*, vol. 58, n.º 7, abr. 2025, doi: 10.1007/s10462-025-11223-9.

[26] T. Hall e K. Rasheed, «A Survey of Machine Learning Methods for Time Series Prediction», *Applied Sciences*, vol. 15, n.º 11, p. 5957, mai. 2025, doi: 10.3390/app15115957.

[27] G. E. P. Box e G. M. Jenkins, *Time series analysis: forecasting and control*, Rev. ed., [Nachdr.]. em Holden-Day series in time series analysis and digital processing. Oakland: Holden-Day, 1979.

[28] Puteri Aiman Syahirah Rosman, Nur Haizum Abd Rahman, Orasa Nunkaw, e Aleta C. Fabregas, «Milestones and developments in classical statistical time series forecasting: a comprehensive review», *Data Anal. Appl. Math.*, pp. 9–22, mar. 2025, doi: 10.15282/daam.v6i1.12113.

[29] P. Domingos, «A few useful things to know about machine learning», *Commun. ACM*, vol. 55, n.º 10, pp. 78–87, out. 2012, doi: 10.1145/2347736.2347755.

[30] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, Adaptive Computation and Machine Learning Series. Cambridge, MA, USA: MIT Press, 2013.

[31] I. Goodfellow, Y. Bengio, e Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[32] J. Brownlee, *Introduction to Time Series Forecasting With Python*, V1.9. Australia: Machine Learning Mastery, 2020. [Online]. Disponível em: https://machinelearningmastery.com/introduction-to-time-series-forecasting-with-python/

[33] B. Benson, W. D. Pan, A. Prasad, G. A. Gary, e Q. Hu, «Forecasting Solar Cycle 25 Using Deep Neural Networks», *Sol Phys*, vol. 295, n.º 5, p. 65, mai. 2020, doi: 10.1007/s11207-020-01634-y.

[34] K. Y. Chan *et al.*, «Deep neural networks in the cloud: Review, applications, challenges and research directions», *Neurocomputing*, vol. 545, p. 126327, ago. 2023, doi: 10.1016/j.neucom.2023.126327.

[35] A. F. Reza, R. Singh, R. K. Verma, A. Singh, Y.-H. Ahn, e S. S. Ray, «An integral and multidimensional review on multi-layer perceptron as an emerging tool in the field of

water treatment and desalination processes», *Desalination*, vol. 586, p. 117849, out. 2024, doi: 10.1016/j.desal.2024.117849.

[36] S. J. D. Prince, *Understanding Deep Learning*. Cambridge, MA, USA: MIT Press, 2024. [Online]. Disponível em: http://udlbook.com

[37] D. P. Kingma e J. Ba, «Adam: A Method for Stochastic Optimization», 30 de janeiro de 2017, *arXiv*: arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980.

[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, e R. Salakhutdinov, «Dropout: A Simple Way to Prevent Neural Networks from Overfitting», *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Disponível em: https://jmlr.org/papers/v15/srivastava14a.html

[39] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, e J. Schmidhuber, «LSTM: A Search Space Odyssey», *IEEE Trans. Neural Netw. Learning Syst.*, vol. 28, n.º 10, pp. 2222–2232, out. 2017, doi: 10.1109/TNNLS.2016.2582924.

[40] F. A. Gers, J. Schmidhuber, e F. Cummins, «Learning to Forget: Continual Prediction with LSTM», IDSIA – Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Lugano, Switzerland, IDSIA-01-99, 1999.

[41] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, vol. 385. em Studies in Computational Intelligence, vol. 385. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-24797-2.

[42] S. Merity, N. S. Keskar, e R. Socher, «Regularizing and Optimizing LSTM Language Models», 7 de agosto de 2017, *arXiv*: arXiv:1708.02182. doi: 10.48550/arXiv.1708.02182.

[43] R. Dey e F. M. Salem, «Gate-variants of Gated Recurrent Unit (GRU) neural networks», em *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Boston, MA: IEEE, ago. 2017, pp. 1597–1600. doi: 10.1109/MWSCAS.2017.8053243.

[44] J. Chung, C. Gulcehre, K. Cho, e Y. Bengio, «Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling», 11 de dezembro de 2014, *arXiv*: arXiv:1412.3555. doi: 10.48550/arXiv.1412.3555.

[45] Z. Jiang, Y. Wang, H. Ning, e Y. Yang, «Research on Application of Convolutional Gated Recurrent Unit Combined with Attention Mechanism in Water Supply Pipeline Leakage Identification and Location Method», *Water*, vol. 17, n.º 4, p. 575, fev. 2025, doi: 10.3390/w17040575.

[46] K. Cho *et al.*, «Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation», em *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1724–1734. doi: 10.3115/v1/D14-1179.

[47] B. N. Oreshkin, D. Carpov, N. Chapados, e Y. Bengio, «N-BEATS: Neural basis expansion analysis for interpretable time series forecasting», 20 de fevereiro de 2020, *arXiv*: airXiv:1905.10437. doi: 10.48550/arXiv.1905.10437.

[48] B. Puszkarski, K. Hryniów, e G. Sarwas, «Comparison of neural basis expansion analysis for interpretable time series (N-BEATS) and recurrent neural networks for heart dysfunction classification», *Physiol. Meas.*, vol. 43, n.º 6, p. 064006, jun. 2022, doi: 10.1088/1361-6579/ac6e55.

[49] C. Challu, K. G. Olivares, B. N. Oreshkin, F. Garza, M. Mergenthaler-Canseco, e A. Dubrawski, «N-HiTS: Neural Hierarchical Interpolation for Time Series Forecasting», apresentado na International Conference on Learning Representations (ICLR), OpenReview, 2023. doi: 10.1609/aaai.v37i6.25854.

[50] J. H. Friedman, «Greedy function approximation: A gradient boosting machine.», *Ann. Statist.*, vol. 29, n.º 5, out. 2001, doi: 10.1214/aos/1013203451.

[51] T. Chen e C. Guestrin, «XGBoost: A Scalable Tree Boosting System», em *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, ago. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[52] R. Shwartz-Ziv e A. Armon, «Tabular data: Deep learning is not all you need», *Information Fusion*, vol. 81, pp. 84–90, mai. 2022, doi: 10.1016/j.inffus.2021.11.011.

[53] G. Ke *et al.*, «LightGBM: A Highly Efficient Gradient Boosting Decision Tree», *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, 2017. [Online]. Disponível em: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

[54] R. J. Hyndman e A. B. Koehler, «Another look at measures of forecast accuracy», *International Journal of Forecasting*, vol. 22, n.º 4, pp. 679–688, out. 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[55] D. Chicco, M. J. Warrens, e G. Jurman, «The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation», *PeerJ Computer Science*, vol. 7, p. e623, jul. 2021, doi: 10.7717/peerj-cs.623.

[56] J. A. Zancanaro, O. L. Dos Santos, L. F. Ugarte, M. Giesbrecht, e M. C. De Almeida, «Energy Consumption Forecasting Using SARIMA and NARNET: An Actual Case Study at University Campus», em *2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, Gramado, Brazil: IEEE, set. 2019, pp. 1–6. doi: 10.1109/ISGT-LA.2019.8895323.

[57] Y. Kim, H. Son, e S. Kim, «Short term electricity load forecasting for institutional buildings», *Energy Reports*, vol. 5, pp. 1270–1280, nov. 2019, doi: 10.1016/j.egyr.2019.08.086.

[58] A. Nanjar, R. E. Saputro, e B. Berlilana, «Machine Learning and Deep Learning Approaches for Energy Prediction: A Systematic Literature Review», *SinkrOn*, vol. 8, n.º 4, pp. 2603–2614, nov. 2024, doi: 10.33395/sinkron.v8i4.14208.

[59] C. Fan, F. Xiao, C. Yan, C. Liu, Z. Li, e J. Wang, «A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning», *Applied Energy*, vol. 235, pp. 1551–1560, fev. 2019, doi: 10.1016/j.apenergy.2018.11.081.

[60] C. Deb, F. Zhang, J. Yang, S. E. Lee, e K. W. Shah, «A review on time series forecasting techniques for building energy consumption», *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902–924, jul. 2017, doi: 10.1016/j.rser.2017.02.085.

[61] M. D. C. Ruiz-Abellón, A. Gabaldón, e A. Guillamón, «Load Forecasting for a Campus University Using Ensemble Methods Based on Regression Trees», *Energies*, vol. 11, n.º 8, p. 2038, ago. 2018, doi: 10.3390/en11082038.

[62] S. A. Alsamraee e S. Khanna, «High-resolution energy consumption forecasting of a university campus power plant based on advanced machine learning techniques», *Energy Strategy Reviews*, vol. 60, p. 101769, jul. 2025, doi: 10.1016/j.esr.2025.101769.

[63] J. Massana, C. Pous, L. Burgas, J. Melendez, e J. Colomer, «Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes», *Energy and Buildings*, vol. 130, pp. 519–531, out. 2016, doi: 10.1016/j.enbuild.2016.08.081.

[64] K. Elhabyb, A. Baina, M. Bellafkih, e A. F. Deifalla, «Machine Learning Algorithms for Predicting Energy Consumption in Educational Buildings», *International Journal of Energy Research*, vol. 2024, n.º 1, p. 6812425, jan. 2024, doi: 10.1155/2024/6812425.

[65] H. Ren, Q. Li, Q. Wu, C. Zhang, Z. Dou, e J. Chen, «Joint forecasting of multi-energy loads for a university based on copula theory and improved LSTM network», *Energy Reports*, vol. 8, pp. 605–612, nov. 2022, doi: 10.1016/j.egyr.2022.05.208.

[66] S. O. Oyinlola e P. O. Oluseyi, «Machine Learning for Campus Energy Resilience: Clustering and Time-Series Forecasting in Intelligent Load Shedding», *arXiv*, set. 2025, doi: 10.48550/arXiv.2509.17097.

[67] A. Vaswani *et al.*, «Attention Is All You Need», 2 de agosto de 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762 .

[68] A. K. Shaikh, A. Nazir, I. Khan, e A. S. Shah, «Short term energy consumption forecasting using neural basis expansion analysis for interpretable time series», *Sci Rep*, vol. 12, n.º 1, p. 22562, dez. 2022, doi: 10.1038/s41598-022-26499-y.

[69] V. Chandola, A. Banerjee, e V. Kumar, «Anomaly detection: A survey», *ACM Comput. Surv.*, vol. 41, n.º 3, pp. 1–58, jul. 2009, doi: 10.1145/1541880.1541882.

[70] C. C. Aggarwal, *Outlier Analysis*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-47578-3.

[71] J. Yang, Q. Song, L. Hu, M. Xin, e R. Xiao, «Power Consumption Anomaly Detection of Smart Grid Based on CAE-GRU», *Energies*, vol. 18, n.º 18, p. 4787, set. 2025, doi: 10.3390/en18184787.

[72] M. Ravinder e V. Kulkarni, «Smart Grid Anomaly Detection Using MFDA and Dilated GRU-based Neural Networks», *Smart Grids and Energy*, vol. 10, n.º 1, p. 9, jan. 2025, doi: 10.1007/s40866-024-00216-2.

[73] D. Sorensen, B. Dey, M. Hwang, e S. Halder, «Unsupervised Anomaly Prediction with N-BEATS and Graph Neural Network in Multi-variate Semiconductor Process Time Series», 23 de outubro de 2025, *arXiv*: arXiv:2510.20718. doi: 10.48550/arXiv.2510.20718.

Anexos

Anexo A

Neste anexo são apresentados os resultados detalhados do processo de otimização de hiperparâmetros realizados com o *Optuna* para os modelos utilizados neste trabalho, incluindo os valores testados e o desempenho obtido em cada ensaio.

Tabela A1: Intervalo de valores testados e melhores hiperparâmetros obtidos durante o processo de otimização do modelo MLP

MLP			
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
MLP (1 camada densa + 1 Dropout)	Learning Rate	1e-5 a 1e-2 (log scale)	0,00105964
	Batch Size	[16, 32, 64, 128]	64
	Nº unidades camada densa	10 a 128 (step=10)	120
	Dropout	0.0 a 0.5 (step=0.05)	0.0
	Função de Ativação	ReLU	ReLU
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0,069076398**
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
MLP (1 camada densa + 1 Dropout) - Com mais Trials (200)	Learning Rate	1e-5 a 1e-2 (log scale)	0.00010523400602435502
	Batch Size	[16, 32, 64, 128]	32
	Nº unidades camada densa	10 a 128 (step=10)	120
	Dropout	0.0 a 0.5 (step=0.05)	0.0
	Função de Ativação	ReLU	ReLU
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0.06791977260073731**
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
MLP (2 camada densa + 1 Dropout)	Learning Rate	1e-5 a 1e-2 (log scale)	0.000403699952560632
	Batch Size	[16, 32, 64, 128]	16
	Nº unidades camada densa (1)	16 a 128 (step=10)	80
	Nº unidades camada densa (2)	16 a 128 (step=10)	112
	Dropout	0.0 a 0.5 (step=0.05)	0.0
	Função de Ativação	ReLU	ReLU
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0,061698741**
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
	Learning Rate	1e-5 a 1e-2 (log scale)	0.0007818688474602626
	Batch Size	[16, 32, 64, 128]	128

MLP (2 camada densa + 1 Dropout)	Nº unidades camada densa (1)	16 a 128 (step=10)	96
	Nº unidades camada densa (2)	16 a 128 (step=10)	64
	Dropout (1)	0.0 a 0.5 (step=0.05)	0.00
	Dropout (2)	0.0 a 0.5 (step=0.05)	0.00
	Função de Ativação	ReLU	ReLU
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0,064268535**

Tabela A2: Intervalo de valores testados e melhores hiperparâmetros obtidos durante o processo de otimização do modelo LSTM

LSTM			
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
LSTM (1 camada densa + 1 Dropout)	Learning Rate	1e-5 a 1e-2 (log scale)	0.0029887038464976924
	Batch Size	[16, 32, 64, 128]	16
	Nº unidades camada densa	16 a 128 (step=16)	96
	Dropout	0.0 a 0.5 (step=0.05)	0.05
	Função de Ativação	tanh	tanh
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0,058209104**

	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
LSTM (2 camada densa + 1 Dropout) - 50 trials	Learning Rate	1e-5 a 1e-2 (log scale)	0.004640413667792685
	Batch Size	[16, 32, 64, 128]	32
	Nº unidades camada densa (1)	16 a 128 (step=16)	128
	Nº unidades camada densa (2)	16 a 128 (step=16)	112
	Dropout	0.0 a 0.5 (step=0.05)	0.1
	Função de Ativação	tanh	tanh
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0,057572417**

	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
LSTM (2 camada densa + 2 Dropout) - 50 trials	Learning Rate	1e-5 a 1e-2 (log scale)	0.0015289653799313884
	Batch Size	[16, 32, 64, 128]	32
	Nº unidades camada densa (1)	16 a 128 (step=16)	112
	Nº unidades camada densa (2)	16 a 128 (step=16)	112
	Dropout (1)	0.0 a 0.5 (step=0.05)	0.3
	Dropout (2)	0.0 a 0.5 (step=0.05)	0.2
	Função de Ativação	tanh	tanh
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0,057110951**

Tabela A3: Intervalo de valores testados e melhores hiperparâmetros obtidos durante o processo de otimização do modelo GRU

Gru			
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
GRU (1 camada densa + 1 Dropout)	Learning Rate	1e-5 a 1e-2 (log scale)	0.0021283708965546644
	Batch Size	[16, 32, 64, 128]	32
	Nº unidades camada densa	16 a 128 (step=16)	112
	Dropout	0.0 a 0.5 (step=0.05)	0.0
	Função de Ativação	tanh	tanh
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0.06037809041441076**

	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
GRU (2 camada densa + 1 Dropout) - 50 trials	Learning Rate	1e-5 a 1e-2 (log scale)	0.0009098352172019962
	Batch Size	[16, 32, 64, 128]	16
	Nº unidades camada densa (1)	16 a 128 (step=16)	48
	Nº unidades camada densa (2)	16 a 128 (step=16)	80
	Dropout	0.0 a 0.5 (step=0.05)	0.5
	Função de Ativação	tanh	tanh
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0.05783706365616067**

	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
GRU (2 camada densa + 2 Dropout) - 50 trials	Learning Rate	1e-5 a 1e-2 (log scale)	0.002733303377179338
	Batch Size	[16, 32, 64, 128]	128
	Nº unidades camada densa (1)	16 a 128 (step=16)	128
	Nº unidades camada densa (2)	16 a 128 (step=16)	80
	Dropout (1)	0.0 a 0.5 (step=0.05)	0.35
	Dropout (2)	0.0 a 0.5 (step=0.05)	0.0
	Função de Ativação	tanh	tanh
	Função de Custo (Loss)	MSE	MSE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0.05715706051353105**

Tabela A4: Intervalo de valores testados e melhores hiperparâmetros obtidos durante o processo de otimização do modelo N-Beats

N-Beats			
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
N-BEATS (input livre)	Learning Rate	1e-5 a 3e-3 (log scale)	0.000934
	Batch Size	[16, 32, 64, 128]	128
	Input Size (janela)	[12, 48]	48
	Função de Normalização (Scaler)	['identity', 'standard', 'minmax']	identity
	Função de Custo (Loss)	MAE	MAE

	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0.05846**

	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
N-BEATS (input fixo = 24 h)	Learning Rate	1e-5 a 3e-3 (log scale)	0.002873
	Batch Size	[16, 32, 64, 128]	16
	Input Size (janela)	Fixo = 24 h	24
	Função de Normalização (Scaler)	['identity', 'standard', 'minmax']	identity
	Função de Custo (Loss)	MAE	MAE
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	0.06182

Tabela A5: Intervalo de valores testados e melhores hiperparâmetros obtidos durante o processo de otimização do modelo XGBoost

XGBoost			
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
XGBoost (Optuna 1)	n_estimators	100 a 1000	903
	max_depth	3 a 10	7
	learning_rate	0.01 a 0.3 (log scale)	0.01894
	subsample	0.5 a 1.0	0.81491
	colsample_bytree	0.5 a 1.0	0.94579
	gamma	0 a 5	0.00554
	Função de Custo (Loss)	reg:squarederror	reg:squarederror
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0,0673**

	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
XGBoost (Optuna 2 – mais parâmetros)	n_estimators	300 a 1200	979
	max_depth	4 a 12	10
	learning_rate	0.005 a 0.2 (log scale)	0.01109
	subsample	0.6 a 1.0	0.73796
	colsample_bytree	0.6 a 1.0	0.95823
	gamma	0 a 5	0.00359
	min_child_weight	1 a 10	10
	reg_alpha	0.0 a 1.0	0.21136
	reg_lambda	0.0 a 5.0	0.16808
	Função de Custo (Loss)	reg:squarederror	reg:squarederror
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0,066**

Tabela A6: Intervalo de valores testados e melhores hiperparâmetros obtidos durante o processo de otimização do modelo N-Hits

N Hits			
	Parâmetro	**Valores Testados**	**Melhor Valor Encontrado**
N Hits (Optuna 1)	n_estimators	100 a 1000	903
	max_depth	3 a 10	7
	learning_rate	0.01 a 0.3 (log scale)	0.01894
	subsample	0.5 a 1.0	0.81491
	colsample_bytree	0.5 a 1.0	0.94579
	gamma	0 a 5	0.00554
	Função de Custo (Loss)	reg:squarederror	reg:squarederror
	Métrica de Avaliação	MAPE	MAPE
	Melhor MAPE obtido	**-**	**0.0673**