

Previsão de Séries Temporais Aplicadas a Análise de Carga de Bateria

Israel R. Dutra

¹Instituto Federal de Educação, Ciências e Tecnologia de São Paulo (IFSP-GRU)

israelr.dtr@gmail.com

Abstract. This work presents the development and analysis of a time series forecasting model applied to the battery current of the GOCE satellite. Using a dataset containing 327 parameters collected throughout the mission, the PHT10040 time series, corresponding to the main battery current, was selected. Raw and filtered data (10 minutes and 6 hours) were used, with processing to remove and replace missing values. The XGBoost model was employed with multiple temporal features and lags, validated via TimeSeriesSplit with five divisions. The results demonstrate a low error rate and forecasts close to the actual values, highlighting the model's efficiency in capturing seasonal patterns and the satellite's energy consumption characteristics.

Resumo. Este trabalho apresenta o desenvolvimento e a análise de um modelo de previsão de séries temporais aplicado à corrente de carga da bateria do satélite GOCE. Com um conjunto de dados contendo 327 parâmetros coletados durante toda a missão, foi selecionada a série temporal PHT10040, referente à corrente da bateria principal. Foram utilizados dados brutos e dados filtrados (10 minutos e 6 horas), com tratamento para remoção e substituição de valores ausentes. O modelo XGBoost foi empregado com múltiplas features temporais e lags, validados via TimeSeriesSplit com cinco divisões. Os resultados demonstram baixa taxa de erro e previsões próximas aos valores reais, evidenciando a eficiência do modelo na captura dos padrões sazonais e das características de consumo energético do satélite.

1. Introdução

O satélite GOCE (*Gravity field and steady-state Ocean Circulation Explorer*) foi lançado pela ESA (*European Space Agency*) em 2009 com a missão de mapear o campo gravitacional da Terra com alta precisão. Durante a sua missão o satélite coletou dados de 327 séries temporais por um período de março de 2009 até outubro de 2013. Para essa aplicação foi usada uma série temporal de corrente de carga da bateria para prever qual será a carga da bateria com base nos dados reais coletados pelo satélite. Foi aplicado um modelo de regressão de séries temporais XGBOOST juntamente com aplicações de hiperparâmetros e validação cruzada.

2. Objetivo

Desenvolver um modelo de machine learning capaz de prever a evolução temporal da carga da bateria.

3. Justificativa

A bateria é a parte principal de um dos sistemas críticos de um satélite, o EPS (*Electrical Power System*) tem como função gerenciar a distribuição de energia da bateria para todo o satélite. Analisar a carga e descarga de corrente do satélite é fundamental para caracterizar o consumo de potência futuro e identificar possíveis falhas sistemáticas do satélite.

4. Metodologia

4.1. Visualizando os dados

Dentre os 327 parâmetros foi selecionado o parâmetro PHT10040. Esse parâmetro contém os dados temporais de carga de corrente da bateria principal do satélite. Dentro dele existem três tipos de dados no formato **.parquet**:

- Dados brutos de toda série temporal. **PHT10040 raw**
- Dados filtrados a cada 10 minutos realizando uma média dos dados brutos. **PHT10040 STATS 10min**
- Dados filtrados a cada 6 horas realizando as médias dos dados brutos. **PHT10040 STATS 6h**.

Para as séries temporais PHT10040 STATS 10min e PHT10040 STATS 6h foram utilizados apenas os dados relativos aos valores médios de corrente, esses dados estão presentes na coluna *value mean* do *dataframe*. A figura 1 mostra os três tipos de dados é notável a diferença de amplitude de cada sinal por efeito da aplicação do filtro de média.

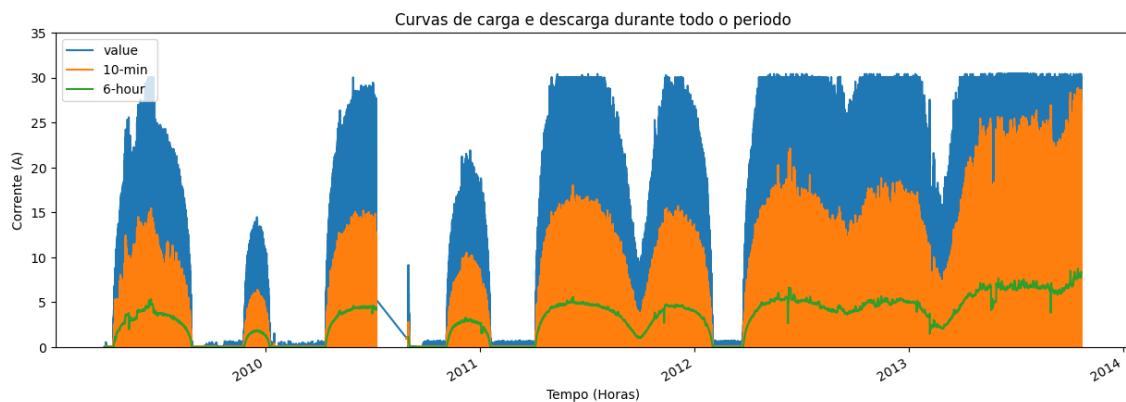


Figura 1. Dados temporais analisados

Além da passagem temporal anual é possível analisar a passagem de tempo para um único dia como é descrito na Figura 2.

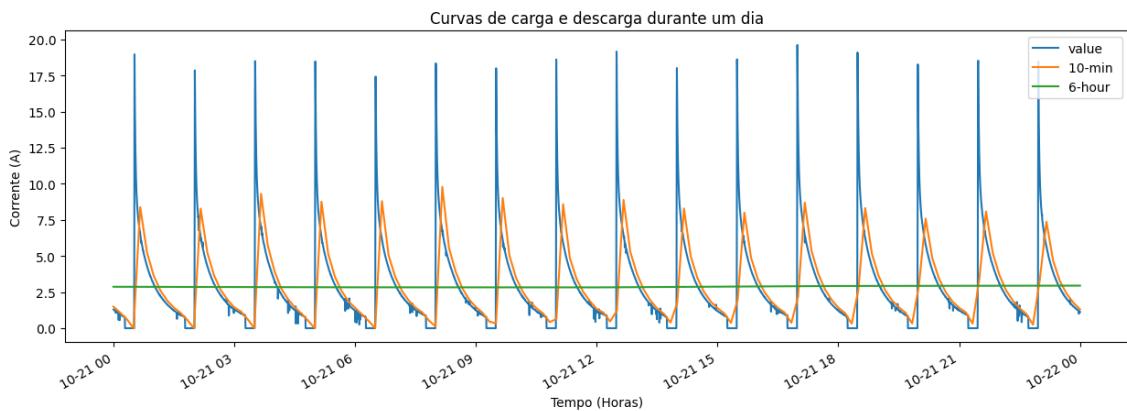


Figura 2. Carga e descarga em um dia

4.2. Tratamento dos dados

Para melhorar a qualidade dos dados foi feita a filtragem dos valores **NAN** para cada serie de dados. Para os dados brutos não foi necessário realizar a filtragem, pois os dados não apresentam nenhum valor vazio. Para os dados de 10 minutos foram 8422 dados vazios e para a serie temporal de 6 horas foram 231 dados vazios. Para contornar esse problema e não apagar nenhum dado foi feita a media dos dias das respectivas series e substituídas nos valores **NAN**.

4.3. Ajustes de hiperparâmetros

Foram implementados cinco parâmetros para o modelo, são eles :

- *Hour*
- *Month*
- *Year*
- *dayofyear*
- *dayofmonth*
- *dayofweek*
- *weekofyear*
- *quarter*

Esses parâmetros servem para caracterizar quais *features* são as mais importante para o modelo. Elas também auxiliam na verificação e validação do modelo, pois auxiliam na criação de arvores de decisão usadas pelo XGBoost criando padrões de sazonalidade, padrões repetitivos e previsão de mudanças.

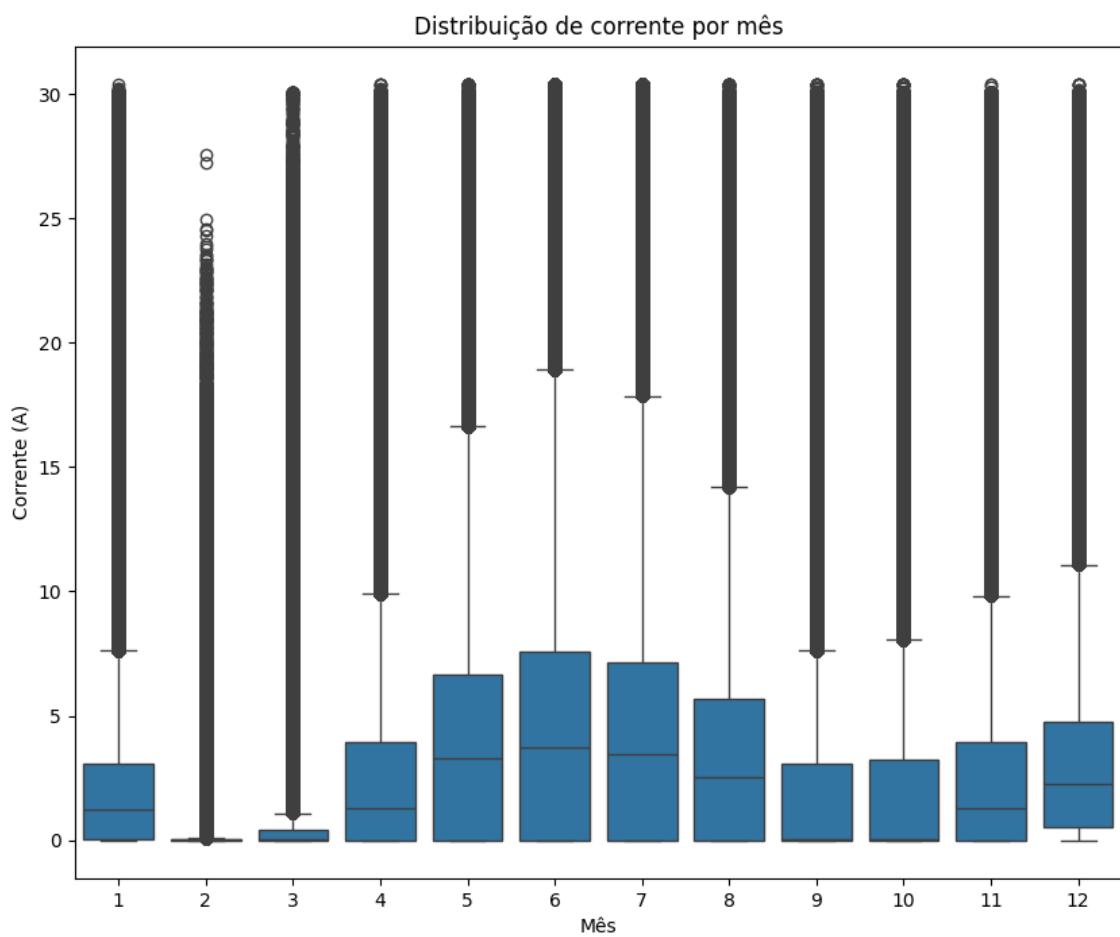


Figura 3. Parâmetro Mês para toda a serie temporal

Também foram aplicadas algumas *features* de *lag*. Esses parâmetros servem como uma memória para o modelo armazenando diversos dados temporais em diferentes períodos de tempo.

4.4. Validação cruzada da serie temporal

Foi implementada a validação cruzada para series temporais utilizando a função *TimeSeriesSplit* da biblioteca *sklearn*. Foram realizados 5 *splits* para cada tipo de serie temporal. Essa etapa tem como função separar de uma maneira mais uniforme e sem bias os dados de treinamento do modelo, assim auxiliando na melhor performance e evitando o *overfitting*.

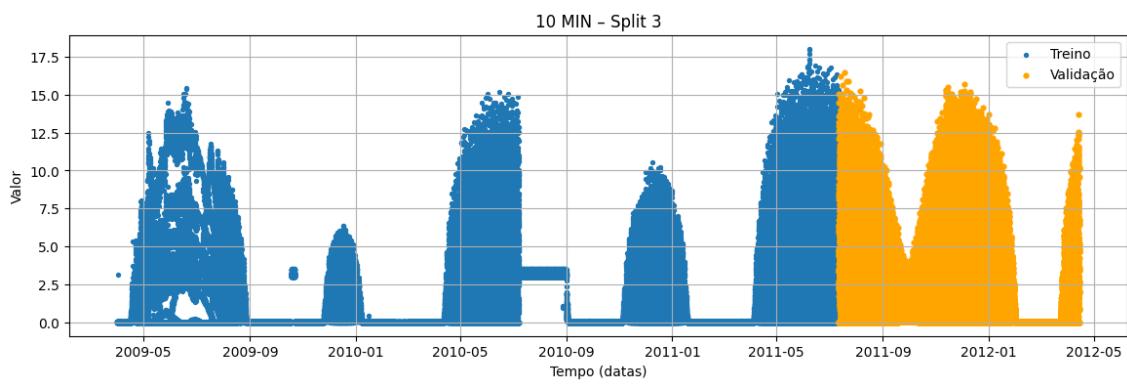


Figura 4. Terceira divisão dos dados de 10 minutos

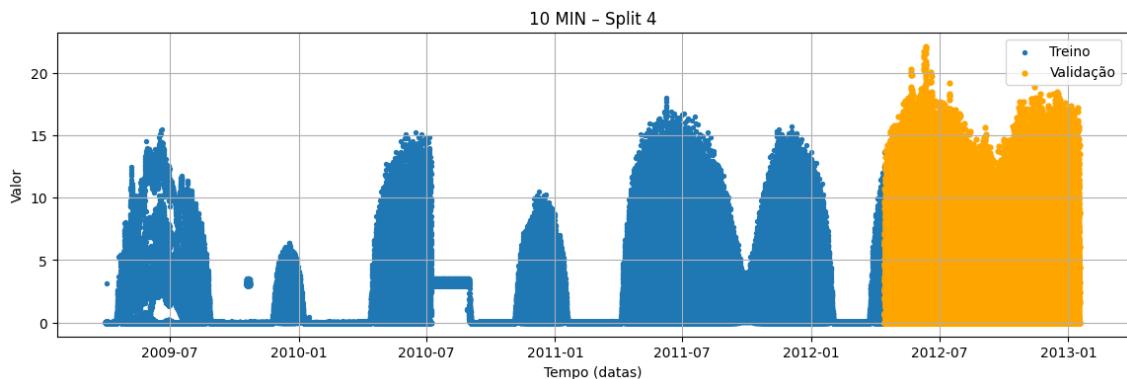


Figura 5. Quarta divisão dos dados de 10 minutos

4.5. Treinamento do modelo

Para realizar o treinamento foi usado o XGBoost

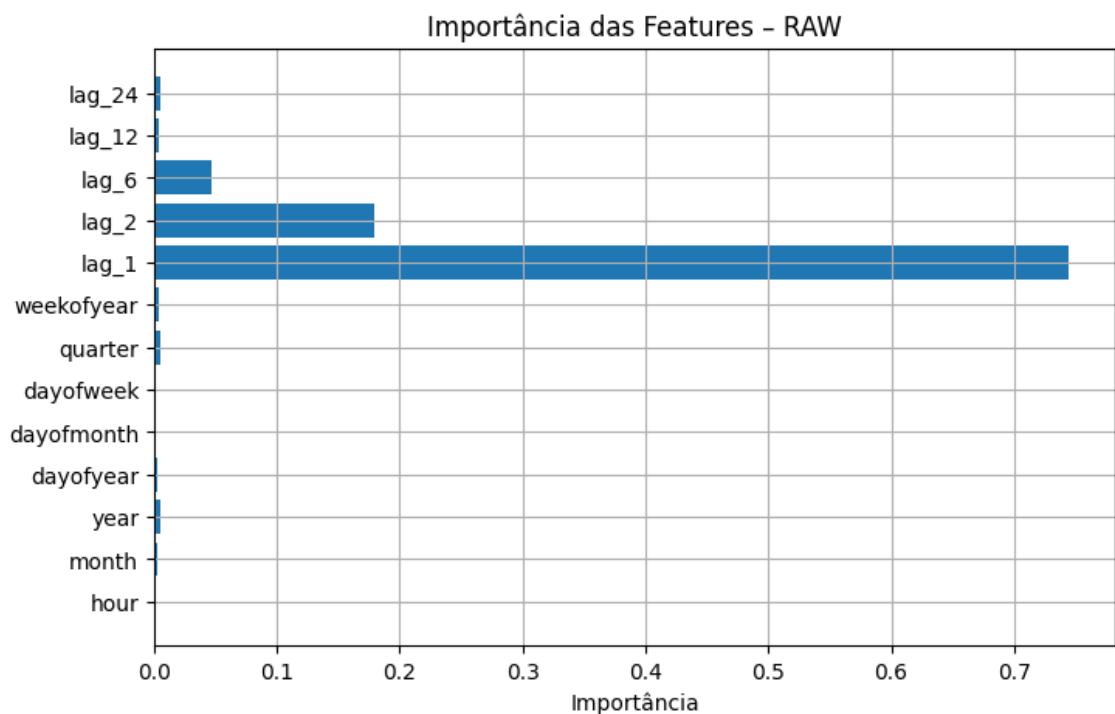


Figura 6. Importância de cada parâmetro para o XGBoost

Como observado na figura acima as features mais importantes são as features de lag. Essas features são as que trabalham como a memória do modelo e portanto afetam drasticamente no seu desempenho.

Dados Brutos		
Split	MAE	RMSE
1	0,065	0,5539
2	0,1206	0,8177
3	0,0913	0,5395
4	0,1944	1,1379
5	0,4338	1,3734

Figura 7. Valores das métricas para os dados brutos

10 minutos		
Split	MAE	RMSE
1	0,196	0,4805
2	0,2836	0,6646
3	0,4407	1,0501
4	0,426	0,8285
5	0,7863	1,4998

Figura 8. Valores das métricas para os dados de 10 minutos

6 horas		
Split	MAE	RMSE
1	0,0764	0,1441
2	0,1751	0,3498
3	0,111	0,1968
4	0,1321	0,2418
5	0,4781	0,6821

Figura 9. Valores das métricas para os dados de 6 horas

As figuras 7, 8 e 9 representam as métricas adquiridas após o treinamento. É possível observar que os valores de *MEAN ABSOLUTE ERROR* e de *ROOT MEAN SQUARE ERROR* permaneceram baixos para todos as séries temporais.

5. Resultados

5.1. Previsões

Para cada tipo de dado foi feita uma previsão com os dados da série temporal e os dados do modelo já treinado, os resultados estão descritos abaixo:

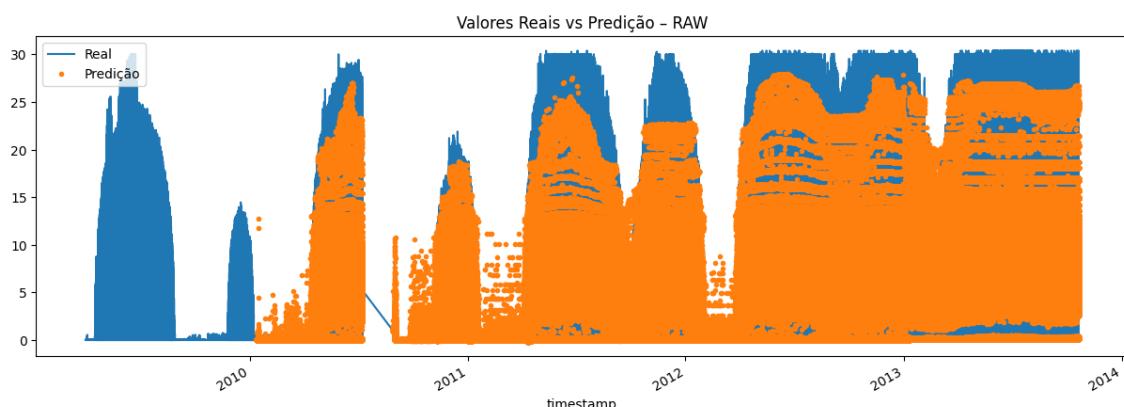


Figura 10. Previsão dos dados brutos

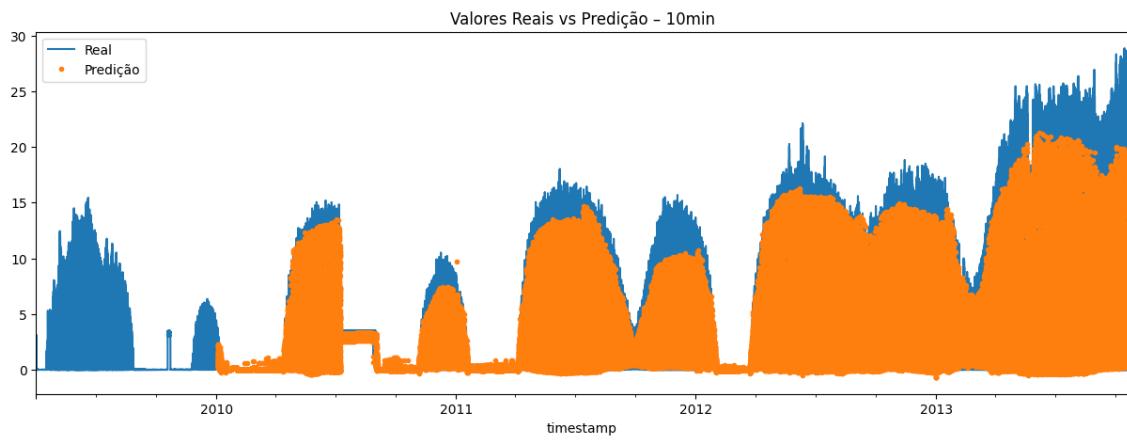


Figura 11. Previsão dos dados de 10 minutos

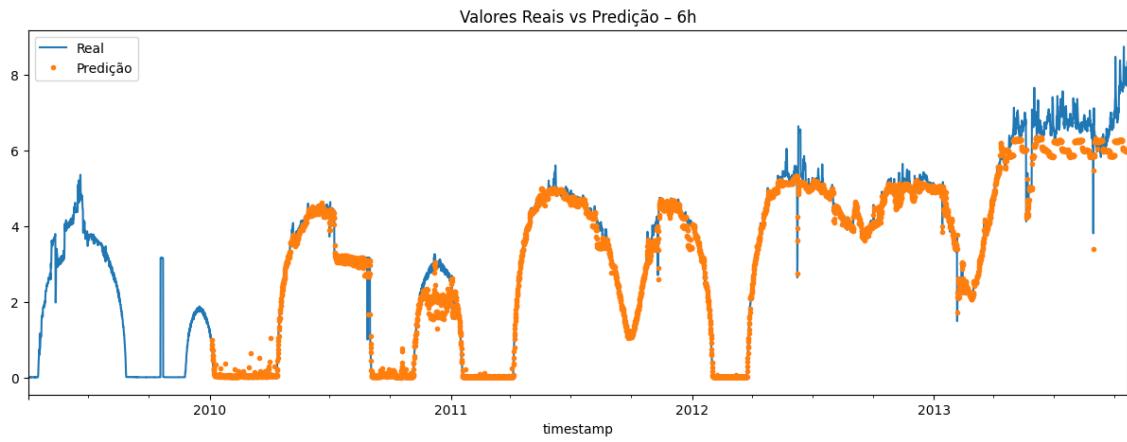


Figura 12. Previsão dos dados de 6 horas

6. Discussão

1. Com base nos dados é possível observar uma alta no consumo de corrente nos últimos anos da missão. Esse fato afeta diretamente o modelo, pois ataca diretamente a sazonalidade dos dados deixando assim as previsões com um maior erro.
2. Conforme a Figura 2 mostra é possível ver as 16 vezes em que o GOCE orbita a terra. Esse fenômeno causa as curvas de carga e descarga presente na figura.
3. Existe uma grande diferença na qualidade dos dados quando comparamos os dados brutos com os dados de 10min e com os dados de 6h. Isso é causado pelo filtro de média que foi aplicado nas series temporais causando amortização dos dados.
4. Uma aplicação de forecasting multivariado poderá ser aplicada no futuro para melhorar a caracterização de consumo de bateria. Apenas um parâmetro analisado pode não oferecer um diagnóstico completo.

5. Como podemos ver na figura 6 o parâmetro que surtiu mais efeito foi o lag1 que corresponde a um passo atrás. Isso significa que gravar o passo anterior para esse caso é o dado mais valioso.
6. Como observado nas imagens 7,8 e 9 o modelo obteve métricas de erros muito baixas. Isso indica que houve pouca ação de outliers no modelo.
7. De modo geral as previsões foram bem próximas dos dados reais. Uma das características que será necessário melhorar em trabalhos futuros é o aumento da amplitude dos valores de corrente.
8. Visando trabalhos futuros, se põe necessária a implementação de uma previsão fora dos dados presentes nos datasets.

7. Conclusão

O modelo demonstrou um bom desempenho para fazer previsões de series temporais. Com esse algoritmo é possível prever como serão as curvas descarga de bateria para o GOCE. As features de lag mostraram-se essenciais para o bom desempenho do modelo, atuando como memória temporal e capturando padrões de variação da série. A análise também evidenciou diferenças significativas entre dados brutos e dados filtrados, resultado do amortecimento causado pelos filtros de média. Para trabalhos futuros, recomenda-se explorar abordagens multivariadas, prever períodos além do intervalo disponível e aprimorar a modelagem da amplitude dos sinais.

8. Referências

Fleith2024GOCE, author = Patrick Fleith, title = GOCE Satellite Telemetry Dataset, howpublished = Kaggle, year = 2024, note = Accessed on 01 December 2025, url = <https://www.kaggle.com/datasets/patrickfleith/goce-satellite-telemetry>