

# Predição de Tempo de Ciclo utilizando Abordagens de Séries Temporais

*Cycle Time Prediction Using Time Series Approaches*

**Laislla C. P. Brandão<sup>1</sup>**

 [orcid.org/0009-0000-0385-8094](https://orcid.org/0009-0000-0385-8094)

**Afrânio Gonçalves<sup>1</sup>**

 [orcid.org/000-0001-6847-709X](https://orcid.org/000-0001-6847-709X)

**Bruno Dantas<sup>1</sup>**

 [orcid.org/0009-0004-3809-1865](https://orcid.org/0009-0004-3809-1865)

**Tiago Teixeira<sup>1</sup>**

 [orcid.org/0000-0002-2225-519X](https://orcid.org/0000-0002-2225-519X)

**Alexandre M. A. Maciel<sup>1</sup>**

 [orcid.org/0000-0003-4348-9291](https://orcid.org/0000-0003-4348-9291)

<sup>1</sup>Escola Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil. E-mail: [lcpb@ecomp.poli.br](mailto:lcpb@ecomp.poli.br)

**DOI: 10.25286/rep.v9i1.2778**

Esta obra apresenta Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Como citar este artigo pela NBR 6023/2018: Laislla C. P. Brandão; Afrânio Gonçalves; Bruno Dantas; Tiago Teixeira. Predição de Tempo de Ciclo utilizando Abordagens de Séries Temporais. Revista de Engenharia e Pesquisa Aplicada, v.9, n. 1, p. 50-59, 2024. DOI: 10.25286/rep.v9i1.2778

## RESUMO

O desenvolvimento de um agente de priorização e a análise estatística desempenham papéis fundamentais na busca por otimizar o tempo de ciclo e melhorar a eficiência produtiva, possibilitando antecipar melhor desvios no tempo de ciclo e a levantar oportunidades de intervenção com antecedência. Analisando o problema de tempo de ciclo por uma abordagem de séries temporais, foram desenvolvidos modelos de predição que permitissem prever o valor de tempo de ciclo em um instante futuro. Três abordagens foram implementadas neste estudo, tendo como base os trabalhos relacionados: um modelo ARIMA, um modelo híbrido consistindo de ARIMA + MLP, e uma rede recorrente LSTM. Os resultados sugerem que as três abordagens possuem bons resultados quando aplicados à série de tempo de ciclo e resultados mais precisos são encontrados quando são combinados modelos que modelam os componentes lineares e não lineares das séries temporais.

**PALAVRAS-CHAVE:** Tempo de ciclo; séries temporais, predição.

## ABSTACT

The development of a prioritization agent and statistical analysis play fundamental roles in the search for optimizing cycle time and improve production efficiency, making it possible to better anticipate deviations in cycle time and identify intervention opportunities in advance. Analyzing the cycle time problem by a time series approach, prediction models were developed that would allow predicting the cycle time value at a future instant. Three approaches were implemented in this study, based on related work: an ARIMA model, a hybrid model consisting of ARIMA + MLP, and an LSTM recurrent network. The results suggest that the three approaches have good results when applied to the cycle time series and more accurate results are found when approaches that model the linear and nonlinear components of the time series are combined.

**KEY-WORDS:** Cycle time, time series, prediction.

## 1 INTRODUÇÃO

### 1.1 CONTEXTUALIZAÇÃO

O mercado e a indústria automotiva estão em constante processo de renovação e transformação, sendo reconhecidos como lideranças na aplicação de novos conceitos e tecnologias inovadoras. Com o advento da indústria 4.0, as montadoras de automóveis precisam adaptar-se rapidamente às tendências para manter sua competitividade em um mercado disputado e continuar atingindo suas metas e superando os desafios.

Uma estimativa do tempo de ciclo de um produto em fábricas de diferentes setores é crítica para avaliar as datas de vencimento do cliente, agendar recursos e ações para a conclusão antecipada do trabalho e para monitorar a operação [8].

No setor de funilaria (body shop), onde inicia-se o processo de fabricação dos veículos e possui grande impacto nos tempos de ciclos para entrega aos clientes finais, a eficiência produtiva e seus indicadores devem estar em linha com o mix esperado pelo negócio. Bem como as demais áreas internas de produção, a fabricação da carroceria e partes móveis é formada por diversas ilhas responsáveis por uma determinada parte do processo, a exemplo de estamparia, soldas, aplicadores de compostos de vedação e geometria dos veículos, onde o resultado final depende do desempenho destas.

### 1.2 DESCRIÇÃO DO PROBLEMA

Existe a necessidade de desenvolver um agente de priorização e análise estatística para auxiliar as equipes de Manutenção e Engenharia a rastrear desvios de tempo de ciclo no processo para buscar oportunidades de realizar intervenções com antecedência e melhorar a eficiência. Desta forma, os times podem antever problemas de eficiência e produtividade antes que estes ocorram.

### 1.3 OBJETIVOS

#### 1.3.1 Objetivo Geral

Desenvolver modelos de aprendizado de máquina para os dados de tempo de ciclo da linha de produção associadas a Carrocerias, de forma a identificar e prever o valor de tempo de ciclo em um instante futuro, possibilitando sinalizar tendências de desvios de tempo de ciclo.

#### 1.3.2 Objetivos Específicos

- Coletar amostras de dados do banco de dados de tempo de ciclo para tratamento e análise;
- Realizar o pré-processamento dos dados, prepará-los e modelá-los para que deles sejam extraídas as informações;
- Encontrar padrões nos dados de funcionamento das linhas de produção para gerar as prioridades das atividades de melhoria de tempo de ciclo para as que estiverem fora de seu *target*;
- Modelar o sistema para que seja possível prever valores futuros nos dados de tempo de ciclo das estações;

### 1.4 JUSTIFICATIVA

Realizar uma análise aprofundada dos dados de tempo de ciclo das linhas de produção é importante, uma vez que fornece informações sobre seus padrões de funcionamento e também gargalos intrínsecos do processo. Além dos gargalos já presentes no processo, já que linhas de produção diferentes possuem máquinas diferentes, podem surgir gargalos pontuais caso a análise apresente dados fora dos padrões normais das estações e linhas. Por meio disso, é possível sinalizar aos tomadores de decisão quais linhas, estações e máquinas apresentam um comportamento mais lento ou tem essa tendência, o que influencia ou irá influenciar a eficiência da produção.

### 1.5 ESCOPO NEGATIVO

- O projeto partirá da base de dados de tempo de ciclo de todas as linhas coletada por um coletor OPC, armazenados no Google Big Query.
- Está fora do escopo a análise de dados prévios a 01/01/2023.
- A base de dados utilizada, assim como os códigos utilizados não serão disponibilizados devido à propriedade intelectual da empresa estudada, e apenas os algoritmos e parâmetros selecionados serão informados.

## 2 FUNDAMENTAÇÃO TEÓRICA

A Funilaria da planta automotiva estudada possui milhares de máquinas distribuídas em 33 linhas

produtivas, das quais uma será o objeto de estudo neste trabalho.

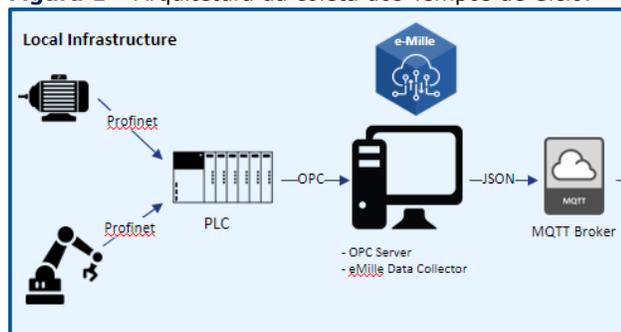
## 2.1 TEMPO DE CICLO FUNILARIA

O tempo de ciclo é definido como o tempo necessário para realizar todas as atividades fabris e agregar valor a uma carroceria, considerando cada elemento da estrutura técnica, como estações e máquinas. Está relacionado à eficiência do processo de produção, pois mais carrocerias podem ser produzidas uma vez que o tempo de ciclo é mais baixo.

Durante a coleta de dados, as estações e máquinas foram modeladas, tendo o tempo de ciclo medido de forma independente para que a união das máquinas de uma estação possa ou não estar contida no tempo da estação individualmente.

As medições de tempo de ciclo são feitas considerando os sinais de quando a carroceria do carro chega na Estação ou Máquina e depois quando sai dela, sendo os valores subtraídos.

**Figura 1** – Arquitetura da coleta dos Tempos de Ciclo.



**Fonte:** Os autores.

## 2.2 MINERAÇÃO DE DADOS

A mineração de dados desempenha um papel fundamental com o objetivo de descobrir padrões e identificar oportunidades de melhoria.

Aplicando técnicas de mineração de dados, como algoritmos de análise de séries temporais ou algoritmos de clusterização, é possível identificar padrões e tendências no tempo de ciclo, bem como identificar correlações entre o tempo de ciclo e outros fatores, como a carga de trabalho, a disponibilidade de recursos ou a configuração das máquinas. Isso pode incluir a detecção de variações sazonais, picos de demanda ou períodos de baixa eficiência. A análise dos padrões ajuda a compreender melhor o desempenho do processo e a identificar áreas que requerem atenção especial.

Uma estimativa do tempo de ciclo de um produto em uma fábrica é fundamental para que os fabricantes avaliem os prazos de entrega dos clientes, programem recursos e ações para a conclusão antecipada do trabalho e monitorem a operação [8]. Os dados históricos podem ser usados para aprender um modelo preditivo para o tempo de ciclo com base em métricas de processo.

## 2.3 SÉRIES TEMPORAIS

Os dados coletados em intervalos regulares de tempo são chamados de dados de série temporal e cada ponto de dados é igualmente espaçado no tempo. A previsão de série temporal é o método de previsão de tendências/padrões futuros de um determinado conjunto de dados históricos com características temporais [2].

Segundo Zang et al. [9], séries temporais podem ser consideradas como composições de padrões lineares e não lineares, logo a combinação de métodos lineares e não lineares pode melhorar a precisão do sistema de previsão da série. O mesmo é afirmado por Babu et al. [1], dizendo que uma combinação adequada de modelos lineares e não lineares fornece um modelo de previsão mais preciso do que um modelo linear ou não linear individual para previsão de dados de séries temporais de várias aplicações.

Uma das estratégias de combinação mais utilizadas e aplicadas a problemas de previsão de séries temporais a fim de aumentar a acurácia dos resultados é a estratégia de hibridizar a série. Nessa metodologia de hibridização, os componentes da série temporal são separados e, então, modelados sequencialmente com a escolha de modelos individuais apropriados [5]. O modelo híbrido com a melhor acurácia é desenvolvido ao determinar o arranjo adequado dos modelos individuais. Portanto, uma das maiores dificuldades ao construir modelos híbridos de séries temporais é como escolher a sequência apropriada de modelos individuais em um procedimento de modelagem sequencial. Em geral, esses sistemas híbridos possuem três etapas principais, que são prever as séries temporais, prever os resíduos (erros) e, então, combinar essas previsões.

Uma vez que balancear o viés e a variância do modelo é uma tarefa desafiadora, os autores em [3] dizem que uma forma de contornar esse desafio que é proposta na literatura é utilizar uma abordagem híbrida, considerando aplicar um modelo linear em um primeiro estágio e um modelo

não linear em um segundo estágio, de forma que a combinação dos dois modelos solucione de forma mais eficaz o problema de previsão.

## 2.4 TRABALHOS RELACIONADOS

Alguns trabalhos que utilizam abordagens híbridas para séries temporais são, por exemplo, o trabalho [4], no qual os autores utilizam um modelo híbrido com ARIMA e SVR para modelar os componentes lineares e não lineares, respectivamente. Aqui, a melhoria destacada é que tanto os parâmetros do ARIMA quanto os do SVR e também os *lags* da série temporal são simultaneamente selecionados usando tanto uma versão contínua quanto uma discreta de um algoritmo de otimização de enxame de partículas (PSO).

Assim como feito em [4], o trabalho [7] também utiliza um modelo ARIMA para modelar os componentes lineares da série em conjunto com um SVR para modelar os não lineares, explorando a força de ambos os modelos em problemas de previsão de séries temporais. Os resíduos obtidos pelo ARIMA são modelados pelo SVM e a ele é acrescentado um erro aleatório. O resíduo final é, então, a soma dos resíduos provenientes do ARIMA e do SVR. Nesse estudo, a previsão de apenas um horizonte de tempo é considerada, pois previne problemas associados a erros acumulados de períodos anteriores. Segundo os autores, hibridizar dois modelos não similares reduz os erros de previsão, mas uma simples combinação dos dois melhores modelos individuais não necessariamente produz os melhores resultados e a seleção dos parâmetros ótimos do modelo híbrido é de grande importância.

Os autores em [1] abordam o uso de sistemas híbridos utilizando ARIMA e ANN, com o foco para um ou múltiplos passos à frente. O diferencial deste trabalho foi primeiramente explorar as naturezas de dados da série, utilizando um filtro de média móvel (MA) e só depois foram aplicados o ARIMA e a ANN para previsão, primeiramente caracterizando os dados fornecidos com base na natureza dos dados. Este filtro decompõe a série em dois modelos de dados distintos, para aplicar a cada modelo uma previsão mais adequada (Linear ou Não Linear). Esta abordagem obteve melhores valores que outras abordagens híbridas, como por exemplo, o trabalho de Zang et al. [9], obtendo melhores resultados tanto em aplicações de um único, como também de múltiplos passos à frente.

A rede LSTM recorrente é outra técnica utilizada e têm capacidade de lidar com as limitações das técnicas tradicionais de previsão de séries temporais, adaptando as não linearidades de um determinado conjunto de dados e podem resultar em resultados de última geração em dados temporais. Cada bloco do LSTM opera em intervalos de tempo diferentes e passa sua saída para o próximo bloco até que o bloco LSTM final gere a saída sequencial [2]. Redes recorrentes com blocos LSTM são algoritmos eficientes para construir um modelo sequencial de séries temporais. O componente fundamental das redes LSTM são os blocos de memória, que foram inventados para lidar com o desaparecimento de gradientes, memorizando parâmetros de rede por longos períodos.

## 3 MATERIAIS E MÉTODOS

### 3.1 DESCRIÇÃO DA BASE DE DADOS

O conjunto de dados utilizado no presente estudo é a base de Tempos e Ciclos da Funilaria. Os dados das linhas principais da Funilaria são coletados de forma automatizada, através de uma aplicação desenvolvida internamente pela empresa, que utiliza um Coletor OPC para agrupar os dados provenientes dos CLPs (Controlador Lógico Programável) e enviá-los para armazenamento Cloud por meio de um protocolo de comunicação MQTT (*Message Queuing Telemetry Transport*).

Após uma carroceria passar por uma máquina ou estação dentro de uma linha de produção, o tempo de saída é subtraído do tempo de entrada e esse dado é registrado como o tempo de ciclo da máquina ou estação.

A base de dados inicial contém dados de seis linhas de produção (duas de assoalhos, duas de laterais e duas da carroceria) e cada registro contém a linha, a estação, a máquina qual o modelo do carro é produzido, o *timestamp*, além de outras informações adicionais. O conjunto de dados utilizado no presente trabalho contém três meses de dados de produção.

A base consiste em uma estrutura tabular, onde cada linha representa uma unidade produzida em uma máquina ou estação de uma linha, e as variáveis incluídas na base de dados bruta são:

- **Shop:** área produtiva da planta.
- **Line:** linha de produção onde registro foi realizado;

- **ST:** identificação da estação onde a carroceria foi processada;
- **Maq:** identificação da máquina utilizada para processar a carroceria;
- **C\_Good:** ciclo de operação de máquina;
- **TCDData:** valor do último tempo de ciclo registrado;
- **Nseq:** ID de Sequência que se repete a cada 16000;
- **ModelCode:** código do modelo;
- **EngineCode:** código do motor;
- **CSALD:** código do Modelo sendo Produzido;
- **Pallet:** ID do Pallet de Transporte Principal (Geopallet, Carrier ou Skid)
- **Cisnum:** CIS (ID único) que identifica a carroceria na história da planta.
- **TIMESTAMP:** registro do *Timestamp* no formato yyyy-MM-dd HH:mm:ss.SSS (UTC).

#	Column	Dtype
0	Shop	object
1	Line	object
2	ST	object
3	Maq	object
4	C_Good	int64
5	TCDData	int64
6	Nseq	float64
7	ModelCode	float64
8	EngineCode	float64
9	CSALD	float64
10	Pallet	float64
11	Cisnum	float64
12	TIMESTAMP	object

Fonte: Os autores.

Outra base de dados incluída no projeto é a de *Targets*, onde são definidos todos os tempos padrões esperados para cada linha.

### 3.2 ANÁLISE DESCRITIVA DOS DADOS

A análise descritiva dos dados é uma etapa essencial na exploração e compreensão do conjunto de dados, que pode ser realizada tanto antes quanto após o pré-processamento dos dados. Ela envolve o uso de técnicas estatísticas e apresentações gráficas para resumir e descrever as principais características dos dados, com posse dessa análise é possível identificar padrões, tendências e informações relevantes nos dados, além de fornecer uma visão geral do conjunto de dados investigado.

A base de dados bruta analisada contém dados de 6 linhas de produção, com 3 milhões 517 mil e 943 linhas e treze colunas. O conjunto de dados está estruturado com diferentes tipos de dados, como *object*, *int64*, *float64*. Na Figura 2 é possível observar cada variável e seus respectivos tipos de dados.

A análise avança verificando os valores ausentes e verifica-se que é necessário um tratamento direcionado a removê-los ou corrigi-los, identificando quais dados são, de fato, dados de processo e quais dados representam erros de registros. A variabilidade dos dados através da coluna *TCDData* após as etapas de tratamento de dados será apresentada posteriormente na seção 3, na Figura 3.

Figura 2 – Variáveis e respectivos tipos de dados.

### 3.3 PRÉ-PROCESSAMENTO

O pré-processamento de dados é uma etapa crítica na busca por soluções que otimizem o tempo de ciclo e melhorem a eficiência produtiva. A partir da limpeza, integração, seleção e transformação de variáveis, detecção de anomalias e padronização, as equipes podem obter insights valiosos que os ajudarão a antecipar problemas, realizar intervenções e manter-se competitivas em um mercado dinâmico e desafiador.

#### 3.3.1 Limpeza e Integração de Dados

Para atacar as perdas de tempo de ciclo e rastrear desvios no processo, é crucial começar pelo processo de limpeza e integração de dados. Diversas fontes de dados, provenientes das diferentes etapas do processo de fabricação, precisam ser coletadas, unificadas e preparadas para análise. Isso envolve a identificação e correção de inconsistências, duplicidades e valores ausentes nos conjuntos de dados.

Os dados de tempo de ciclo medidos em cada uma das máquinas nem sempre refletem o tempo que, de fato, é gasto em uma estação para performar todos os trabalhos e passar a carroceria adiante. A maioria das máquinas realiza apenas um pequeno trabalho e agregando-se esses dados, o resultado não traz informação suficiente. Os dados medidos para a estação como um todo, por sua vez, refletem a realidade do ciclo da estação.

Por esse motivo, apenas os dados de medição das estações como um todo serão considerados como dados úteis para a condução deste trabalho, assim como já é utilizado em estudos prévios em andamento sobre o tempo de ciclo na planta automotiva considerada. Assim, inicialmente, um filtro na coluna *Maq* filtra apenas o que for dados da medição da estação e desconsidera dados individuais das máquinas.

Como processo de tratamento das informações, tanto para a geração do banco de dados quanto da métrica comparativa, foi feita a remoção de colunas que não serão utilizadas no projeto. Algumas são apenas um ID sequencial, como no caso de *C\_Good* e *Nseq*, já outras não possuem dados na base e existem para uma futura coleta, mas hoje não são usadas.

### 3.3.2 Seleção e Transformação de Variáveis

A seleção e transformação de variáveis são etapas críticas para identificar as características mais relevantes ao processo e a seleção adequada de variáveis e a aplicação de técnicas de transformação, como normalização e padronização, podem simplificar a análise e melhorar a precisão das previsões.

Uma vez que a coluna dos *timestamps* era uma coluna de texto, buscou-se transformá-la para *timestamp* e, posteriormente, criar uma nova coluna de Data e uma coluna de Hora.

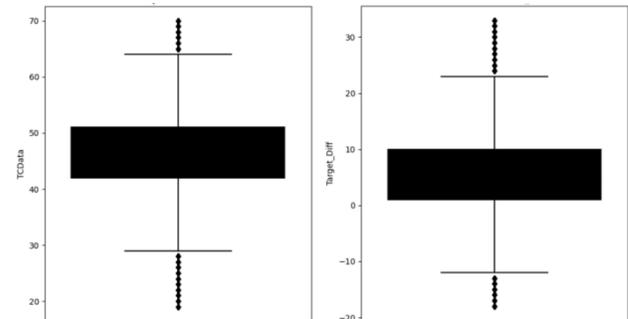
Muitos dos registros de *CSALD* são erros de medição nesta variável e não estão associados a nenhum modelo específico, então foi preciso corrigi-los, alterando estes valores inválidos para zero. Após isto, cada código único foi associado aos nomes de cada modelo específico de carro produzido, sendo criada uma coluna correspondente. Foi alterada a coluna de Estação para nomenclatura correta usando o padrão Linha-Estação e, por fim, feita a inclusão da coluna de *Target* com o valor de tempo de ciclo esperado para cada linha, junto com a criação da coluna de *Target\_Diff*, que é a diferença entre o registro de tempo de ciclo e o *target* da linha.

### 3.3.3 Detecção de Anomalias

A detecção de anomalias é um componente vital do pré-processamento de dados, permitindo a identificação precoce de problemas de eficiência e produtividade. Isso envolve o desenvolvimento de modelos estatísticos e algoritmos de aprendizado de máquina para identificar desvios significativos no processo de fabricação. Em relação aos valores bastante discrepantes dos demais, conhecidos como *outliers*, percebe-se que a base possuía diversos dados na coluna *TCDData*, seja por erros na medição ou porque esses registros ocorreram em períodos que não são verdadeiramente produtivos, como testes e paradas técnicas, por exemplo. Todos os

valores que estivessem acima ou abaixo de um desvio padrão e meio da média foram desconsiderados.

**Figura 3** – Boxplot dos dados de *TCDData* e *Target\_Diff* após remoção de *outliers*.



**Fonte:** Os autores.

Ao final do tratamento e transformação dos dados, não foram encontrados registros faltantes na base, como existiam no conjunto bruto de dados iniciais sem tratamento.

Por fim, após todas as etapas descritas acima, o conjunto utilizado possui 159.650 registros e 11 variáveis, ou *features*, e o conjunto de dados que será utilizado neste projeto está descrito abaixo na Tabela 1, juntamente com o seu dicionário de dados atualizado.

### 3.3.4 Padronização e Normalização

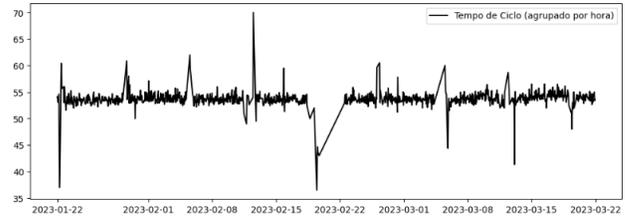
Para garantir a consistência dos dados e facilitar a comparação entre diferentes ilhas de produção e etapas do processo, é essencial padronizar e normalizar os dados. Essas técnicas permitem que os dados sejam comparáveis, independentemente das unidades de medida ou escalas utilizadas. Para manter os dados normalizados em um intervalo entre 0 e 1, foi utilizada a biblioteca *MinMaxScaler*.

**Tabela 1** - Dicionário dos dados de Tempos de Ciclo.

COLUNA	MODO	TIPO	DESCRIÇÃO
<b>Shop</b>	NULLABLE	STRING	Oficina na Planta.
<b>Line</b>	NULLABLE	STRING	Linha na Oficina.
<b>ST</b>	NULLABLE	STRING	Estação na Linha.
<b>Maq</b>	NULLABLE	STRING	Máquina na Estação.
<b>TCDData</b>	NULLABLE	INTEGER	Último tempo de ciclo registrado.

<b>TIMESTAMP</b>	NULLABLE	TIMESTAMP	Registro do Timestamp formato yyyy-MM-dd HH:mm:ss.SSS (UTC).
<b>data</b>	NULLABLE	DATE	Registro da data no formato yyyy-MM-dd.
<b>Modelo</b>	NULLABLE	STRING	Nome do veículo produzido.
<b>descModelo</b>	NULLABLE	STRING	Nome do modelo detalhado do veículo produzido.
<b>TC_Target</b>	NULLABLE	INTEGER	Valor do target de Tempo de Ciclo para a Linha.
<b>Target_Diff</b>	NULLABLE	INTEGER	Diferença entre o valor registrado de tempo ciclo e o target para a linha.

Fonte: Os autores.



Fonte: Os autores.

Para constatar a estacionariedade de uma série, comumente são utilizados testes estatísticos como, por exemplo, o Teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin). Aplicando o teste à série de tempo de ciclo, rejeita-se a hipótese nula, uma vez que a estatística do teste é inferior ao valor crítico a uma significância de 5%. Dessa forma, a série é estacionária.

### 3.4 METODOLOGIA EXPERIMENTAL

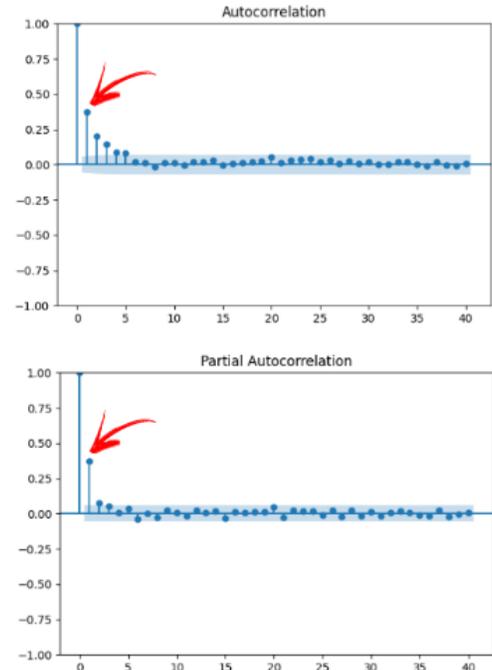
A metodologia visa desenvolver uma análise estatística e um mecanismo de predição para prever o tempo de ciclo e antecipar problemas de eficiência e produtividade. O desenvolvimento de um agente de priorização e a análise estatística desempenham papéis fundamentais na busca por otimizar o tempo de ciclo e melhorar a eficiência produtiva, contribuindo para a competitividade contínua no mercado.

#### 3.4.1 Análise de Séries Temporais

A primeira etapa da análise da série consiste na visualização dos dados através de um gráfico temporal. Este revela as características da série, ou seja, informações sobre tendências, comportamentos sazonais, e outros padrões sistemáticos que possam estar presentes. A partir do conjunto de dados anterior após remoção de outliers, foi filtrada uma única estação produtiva, que será avaliada a partir daqui, e seus dados individuais foram agrupados de hora em hora considerando a média de tempo ciclo no período para criação da série horária. Dessa forma, é possível prever um passo (hora) à frente. A Figura 4 apresenta a série temporal de tempo de ciclo que será utilizada neste estudo.

Figura 4 – Série de dados de tempo de ciclo para uma estação de uma linha produtiva agrupadas a cada hora.

Figura 5 – Autocorrelações e autocorrelações parciais entre os lags da série de tempo de ciclo.



Fonte: Os autores.

A Figura 5 mostra as autocorrelações e autocorrelações parciais entre os lags da série, considerando o conjunto de dados de treino. Estas informações ajudam a definir, respectivamente, a ordem dos modelos de média móvel (MA) e autorregressivo (AR).

#### 3.4.2 Algoritmos de Previsão de Séries Temporais

Três abordagens serão testadas nesse estudo, tendo como base os trabalhos relacionados e buscando a melhor precisão dos resultados para a previsão da série de acidentes. São elas: o modelo ARIMA, um modelo híbrido Linear→Não Linear, consistindo de ARIMA + MLP, e uma rede recorrente LSTM. Para treinamento dos modelos, foram considerados os primeiros 70% dos registros da série. Para o teste, foram separados os 30% dos registros seguintes.

- **ARIMA:** Os dados de treinamento serão utilizados para treinar o modelo e a previsão será feita *step-by-step*, um *rolling* ARIMA. Utilizando o conjunto de teste, será feita a previsão de um passo (hora) à frente pelo ARIMA e o dado do *step* seguinte dos dados de teste vai ser adicionado à série de treinamento para que, assim, seja previsto o horizonte de tempo seguinte. Dessa forma, será realizada uma previsão interativa de um passo à frente. A cada previsão é calculado o resíduo do erro em relação ao valor real do conjunto de teste e serão armazenados em uma série dos resíduos. Foram selecionados parâmetros para o ARIMA de acordo com as autocorrelações, logo  $p = 1$ ,  $d = 0$ ,  $q = 1$ .
- **ARIMA + MLP:** Nesta abordagem, a ideia é utilizar o modelo ARIMA para modelar os componentes lineares da série temporal, seguidos de um modelo MLP para modelar os componentes não lineares dos resíduos da etapa anterior. Assim como em [7], apenas a previsão de um único horizonte de tempo é considerada, a fim de prevenir problemas associados a erros acumulados de períodos anteriores. A série de resíduos é deslocada para criar os *lags* para a rede neural, depois normalizada, modelada pela MLP e a previsão para todos os resíduos é realizada de uma única vez. A partir daí, os dados previstos para os resíduos são somados às previsões provenientes do ARIMA, resultando na previsão geral para o modelo híbrido. Os parâmetros utilizados para o ARIMA são os mesmos e, para a MLP, os parâmetros definidos foram quantidade de *hidden layers* = 3, neurônios das *hidden layers* = (200,150,200), função de ativação = ReLU, otimizador = LBFGS, *learning rate* = 0.001 e quantidade de *lags* para os dados da série = 10.

- **LSTM:** Para esta abordagem, o conjunto de treinamento é deslocado para criar os *lags* para serem inseridos na rede neural, os dados são padronizados utilizando o desvio padrão e a rede é treinada utilizando os dados de teste para validação durante o treinamento. Os parâmetros definidos para a LSTM foram quantidade de camadas LSTM = 2, com 32 e 25 neurônios, respectivamente, cada uma sucedida por uma camada de *dropout* com taxa = 0.25, função de ativação = ReLU, otimizador = Adam, *learning rate* = 0.01 e quantidade de *lags* para os dados da série = 12. Há uma regularização L2 com parâmetro de regularização = 0.01 nas camadas LSTM. A última camada é uma camada densa com 1 neurônio e o modelo é treinado com *batch size* = 25.

### 3.4.3 Métricas de Avaliação

Serão utilizados o Erro Quadrático Médio (*MSE*) e a Raiz do Erro Quadrático Médio (*RMSE*) como métricas de avaliação para todas as abordagens. O *MSE*, por ser uma métrica que apura o quanto alguns resultados passam a se afastar de uma média aguardada inicialmente, e o *RMSE*, por ser o desvio padrão dos valores residuais (erros de previsão) e estar na mesma unidade dos valores utilizados, facilitando a explicabilidade dos resultados para os *stakeholders*.

## 4 RESULTADOS E DISCUSSÕES

Seguindo o protocolo experimental descrito na seção anterior, nesta seção, serão apresentados os resultados dos experimentos realizados, bem como as discussões a respeito deles. Ao todo, foram repetidos dez experimentos para cada uma das abordagens propostas para verificar a consistência dos resultados.

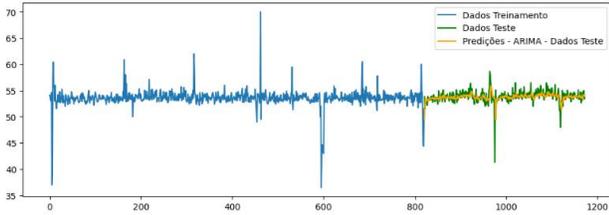
### 4.1 RESULTADOS

O resultado para cada uma das abordagens é apresentado a seguir, refletindo como foram separados os conjuntos de treinamento e teste e como os resultados dos modelos conseguiram refletir o padrão da série em cada caso.

- **ARIMA:** A Figura 6 mostra o resultado das previsões da série para os dados de teste pelo

modelo linear ARIMA para um dos experimentos realizados.

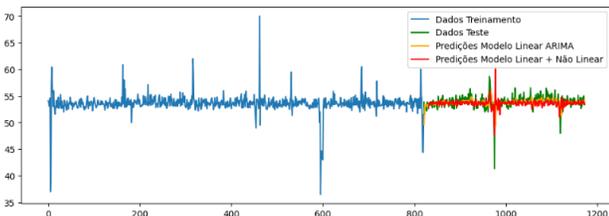
Figura 6 – Resultado da predição - Abordagem ARIMA.



Fonte: Os autores.

- **ARIMA + MLP:** A Figura 7 mostra, não apenas o resultado das previsões da série para o modelo linear ARIMA em um dos experimentos realizados, como também o resultado completo do modelo híbrido ARIMA+MLP, proveniente da composição das previsões dos dados de teste e dos resíduos da série pelo modelo não linear MLP.

Figura 7 – Resultado da predição - Abordagem ARIMA+MLP.

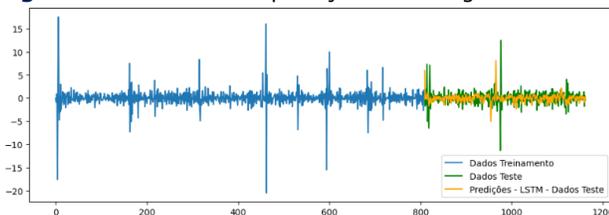


Fonte: Os autores.

- **LSTM:** A seguir, na Figura 8, vê-se o resultado das previsões da série para os dados de teste pelo modelo não linear LSTM para um dos experimentos realizados.

Para os dez experimentos de cada abordagem desenvolvida, na Tabela 2 é possível ver compiladas as médias dos resultados de *MSE* e *RMSE* para cada uma, juntamente com seus respectivos desvios padrões.

Figura 8 – Resultado da predição - Abordagem LSTM.



Fonte: Os autores.

Tabela 2 - Resultado das abordagens aplicadas (média e desvio padrão dos 10 experimentos).

ABORDAGEM APLICADA	MSE	RMSE
ARIMA	1,43 ± 0,00	1,20 ± 0,00
ARIMA + MLP	<b>1,24 ± 0,12</b>	<b>1,11 ± 0,05</b>
LSTM	6,23 ± 0,33	2,50 ± 0,07

Fonte: Os autores.

## 4.2 DISCUSSÃO

Analisando o comportamento da previsão do modelo ARIMA em comparação ao conjunto de teste, percebe-se que o modelo consegue desenhar bem o padrão dos dados, seguindo as variações destes de forma suave. O interessante da comparação na Figura 7, por sua vez, é verificar como o resultado híbrido está levemente mais próximo dos dados de teste do que o do modelo ARIMA unicamente e consegue captar os padrões, sendo capaz de prever inclusive os maiores picos e vales nos dados com bastante precisão. Para o resultado da LSTM, observa-se um resultado parecido, com a reprodução do padrão oscilatório da série, ainda que o modelo não acerte tão bem os maiores picos de dados como o modelo anterior.

É possível perceber que os três modelos apresentam bons resultados para a predição de um passo à frente e o que foi visto qualitativamente ao observar os gráficos com as previsões das diferentes abordagens, mostra-se verdadeiro ao final da comparação quantitativa por meio das métricas. O que se observa no final é um melhor resultado de previsões da série ao utilizar uma abordagem híbrida, permitindo que o modelo resultante seja bom em perceber tanto os padrões lineares quanto os não lineares da série.

O modelo linear ARIMA sozinho ainda consegue captar bem o comportamento linear dos tempos de ciclo, pela própria característica da série, o que mostra que nem sempre um modelo mais complexo necessariamente produzirá os melhores resultados e cada caso precisa ser analisado para a escolha do melhor modelo a ser utilizado.

O resultado da LSTM, por mais que tenha sido o que apresentou as métricas mais altas, ainda está com um bom resultado, considerando que o erro médio de previsão é de apenas ± 2,5 segundos.

Estes resultados precisam ser posteriormente avaliados em um contexto em que os *targets* de tempo de ciclo estejam envolvidos, pois resultados preditos abaixo do *target* não são, a princípio,

senhais de desvios e não requerem a mesma atenção dos times produtivos.

## 5 CONCLUSÕES

Uma análise estatística foi realizada para identificar tendências, correlações e padrões nos dados de tempo de ciclo para possibilitar o levantamento de desvios no tempo de ciclo e a busca de oportunidades de intervenções direcionadas a melhorar a eficiência produtiva.

Analisando o problema como uma abordagem de séries temporais, buscou-se utilizar algumas das técnicas mais utilizadas para definir o modelo de predição com melhor resultado.

O estudo foi realizado para apenas uma das estações de uma linha produtiva de carrocerias, então, para o futuro, é necessário que seja realizado um estudo sobre cada uma das outras estações no que diz respeito à previsibilidade de cada caso, uma vez que nem todas as séries temporais podem ser previstas.

Como trabalhos futuros, é interessante que sejam avaliados outros modelos não lineares, como uma Regressão de Vetor de Suporte, por exemplo, para compor um modelo híbrido que seja capaz de aprender o padrão da série de forma cada vez mais precisa. Além disso, ao utilizar mais de um modelo, é importante pensar na possibilidade de realizar uma seleção dinâmica para definir o melhor modelo para a predição da série, avaliando os menores erros. Além disso, é interessante que seja desenvolvida alguma interface aos usuários para facilitar a informação dos dados previstos e possíveis alertas no caso de valores de tempo ciclo previstos estarem fora de um limite pré-estabelecido acima dos targets.

Por fim, os resultados deste projeto inicial e estudo de tempo de ciclo têm potencial de agregar bastante valor, ajudando a melhorar a eficiência da oficina produtiva e beneficiando times de produção, manutenção e engenharia de manufatura, além de poder ser replicados e expandidos para diversas plantas e setores.

## REFERÊNCIAS

[1] BABU, C. Narendra; REDDY, B. Eswara. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. **Applied Soft Computing**, v. 23, p. 27-38, 2014.

[2] CHIMMULA, Vinay Kumar Reddy; ZHANG, Lei. Time series forecasting of COVID-19 transmission in

Canada using LSTM networks. **Chaos, solitons & fractals**, v. 135, p. 109864, 2020.

[3] DE OLIVEIRA, Joao FL; SILVA, Eraylson G.; DE MATTOS NETO, Paulo SG. A hybrid system based on dynamic selection for time series forecasting. **IEEE Transactions on Neural Networks and Learning Systems**, v. 33, n. 8, p. 3251-3263, 2021.

[4] DE OLIVEIRA, Joao Fausto Lorenzato; LUDERMIR, Teresa Bernarda. A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting. In: 2014 **Brazilian Conference on Intelligent Systems**. IEEE, 2014. p. 73-78.

[5] HAJIRAHIMI, Zahra; KHASHEI, Mehdi. Sequence in hybridization of statistical and intelligent models in time series forecasting. **Neural Processing Letters**, v. 54, n. 5, p. 3619-3639, 2022.

[6] **Hub de conceitos de computação em nuvem** - <https://aws.amazon.com/pt/what-is/>.

[7] PAI, Ping-Feng; LIN, Chih-Sheng. A hybrid ARIMA and support vector machines model in stock price forecasting. **Omega**, v. 33, n. 6, p. 497-505, 2005.

[8] P. Backus, M. Janakiram, S. Mowzoon, C. Runger and A. Bhargava, **Factory cycle-time prediction with a data-mining approach**, in *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 2, pp. 252-258, May 2006, doi: 10.1109/TSM.2006.873400.

[9] ZHANG, G. Peter. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, v. 50, p. 159-175, 2003.