

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Práticas de Aprendizado de Máquina  
Usando Previsão de Consumo de Cimento  
como Estudo de Caso**

Pedro Fernandes

MONOGRAFIA FINAL

MAC 499 — TRABALHO DE  
FORMATURA SUPERVISIONADO

Supervisor: Prof. Dr. Marcelo Finger

São Paulo  
2023

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0  
(Creative Commons Attribution 4.0 International License)*

# Agradecimentos

*Quanto melhor é adquirir a sabedoria do que o ouro!  
E mais excelente, adquirir a prudência do que a prata!*  
— Provérbios 16:16

Agradeço, primeiramente, pela graça de Deus, que me proveu e continua provendo com oportunidades ricas como ter feito esse excelente curso e ter tido capacidade de concluí-lo. Agradeço a minha família, pai, tios e, em especial, minha mãe, que nunca faltou em me apoiar nos momentos mais difíceis.

Agradeço, ao meu orientador, Prof. Dr. Marcelo Finger, pelo excelente equilíbrio entre orientação, espaço para liberdade criativa e apoio moral.

Agradeço a soma de todos os professores, coordenadores e funcionários desse extraordinário instituto por efeturarem um trabalho diferenciado em que é visível a dedicação na transmissão do conhecimento, organização do curso e suporte para todos os seus integrantes.



# Resumo

Pedro Fernandes. **Práticas de Aprendizado de Máquina Usando Previsão de Consumo de Cimento como Estudo de Caso**. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2023.

É de crítica importância a previsão antecipada do consumo de cimento para a indústria produtora e sociedade em geral. Uma previsão assertiva dessa informação beneficiaria diretamente as estratégias de produção, logística, minimização de desperdícios e redução de poluição em todo processo. A precisão dessas previsões, realizadas com um ano de antecedência, baseadas em informações do passado, desempenha um papel vital na orientação das decisões futuras. Neste trabalho, explorou-se a previsão do consumo de cimento por meio de análise de séries temporais e modelos de aprendizado de máquina respeitando boas práticas como regularização entre outras para evitar modelos que sofressem de sobreajuste. Foram utilizados métodos de pré-processamento de dados, como interpolação, transformações e média móvel conjuntamente com a inclusão de novos dados, pertinentes ao domínio do problema, para obter uma melhoria progressiva da assertividade da previsão dos modelos. Observou-se, desta forma, uma redução consistente nos erros e uma maior precisão à medida que novos processamentos e tipos de dados foram introduzidos.

**Palavras-chave:** Aprendizado de Máquina. Séries Temporais. Pré Processamento de Dados. Rede Neural Recorrente. Rede Neural LSTM. Rede Neural Feed-forward. Explicabilidade por valores SHAP. Consumo de Cimento.



# Abstract

Pedro Fernandes. **Machine Learning Practices Using Cement Consumption Forecasting as a Case Study**. Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2023.

The early forecast of cement consumption is of critical importance for the manufacturing industry and society at large. An accurate forecast of this information would directly benefit production strategies, logistics, waste reduction, and pollution reduction throughout the process. The accuracy of these forecasts, made a year in advance based on past information, plays a vital role in guiding future decisions. This study explored cement consumption forecasting through time series analysis and machine learning models, adhering to good practices like regularization to prevent overfitting. Data preprocessing methods such as interpolation, transformations, and moving averages, combined with the inclusion of new domain-relevant data, were used to progressively improve the models' predictive accuracy. Consequently, a consistent reduction in errors and enhanced precision was observed as new processing methods and data types were introduced.

**Keywords:** Machine Learning. Time series. Data Preprocessing. Recurrent Neural Network. LSTM Neural Network. Feed-forward Neural Network. Explainability through SHAP Values. Cement Consumption.



# Lista de abreviaturas

ABNT	Associação Brasileira de Normas Técnicas
ARIMA	Média Móvel Integrada Autoregressiva ( <i>AutoRegressive Integrated Moving Average</i> )
BCB	Banco Central do Brasil
CMA	Média Móvel Centrada ( <i>Centrated Moving Average</i> )
EMBI	Índice de títulos de mercados emergentes ( <i>Emerging Market Bond Index</i> )
FFNN	Rede Neural Feed Forward ( <i>Feed Forward Neural Network</i> )
FGTS	Fundo de Garantia do Tempo de Serviço
FGV	Fundação Getulio Vargas
IA	Norma Brasileira
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
IME	Instituto de Matemática e Estatística
INCC	Índice Nacional de Custo da Construção Civil
INMET	Instituto Nacional de Meteorologia
IPCA	Índice Nacional de Preços ao Consumidor Amplo
IPEA	Instituto de Pesquisa Econômica Aplicada
LSTM	Memória de longo e curto prazo ( <i>Long-Short Term Memory</i> )
MAE	Média Percentual de Erro Absoluto ( <i>Mean Absolute Error</i> )
MAPE	Média Percentual de Erro Absoluto ( <i>Mean Absolute Percentage Error</i> )
ML	Machine Learning
NBR	Norma Brasileira
PIB	Produto Interno Bruto
RNN	Rede Neural Recorrente ( <i>Recurrent Neural Network</i> )
SGD	Gradiente Descendente Estocástico ( <i>Stochastic Gradient Descent</i> )
SNIC	Sindicato Nacional da Indústria do Cimento
RMSE	Média Raiz do Erro Quadrado ( <i>Root Mean Squared Error</i> )
USP	Universidade de São Paulo

## Lista de figuras

2.1	Exemplo de interpolação linear. . . . .	8
2.2	Exemplo de interpolação polinomial. . . . .	9
2.3	Exemplo de interpolação polinomial com dados imputados. . . . .	10
2.4	Exemplo de médias móveis. . . . .	13
2.5	Rede neural feed-forward. Figura de <a href="#">BARRERA-ANIMAS et al. (2022)</a> <sup>1</sup> . . . . .	15
2.6	Superfície da função de custo e exemplo de pontos de mínimo local e global. . . . .	17
2.7	Delimitações da Rede LSTM. Figura de <a href="#">MASINI et al. (2023)</a> <sup>2</sup> . . . . .	18
2.8	Célula LSTM. Figura de <a href="#">MASINI et al. (2023)</a> <sup>3</sup> . . . . .	19
2.9	Exemplo de correção de sobreajuste por regularização. . . . .	24
2.10	Figura explicativa de valores SHAP. . . . .	27
3.1	Exemplo de imputação com sazonalidade média. . . . .	36
3.2	Stacked-LSTM. Figura adaptada de <a href="#">YU et al. (2019)</a> <sup>4</sup> . . . . .	37
3.3	Exemplo média de sazonalidade dos dados de treino. . . . .	39
3.4	Fluxo do treinamento. . . . .	41
4.1	Distribuição agregada dos erros dos dados iniciais brutos. . . . .	45
4.2	Resultados das previsões de cimento para 2019 - Dados iniciais. . . . .	46
4.3	Distribuição agregada dos erros dos dados iniciais com pré-processamento adicional. . . . .	48
4.4	Resultados das previsões de cimento para 2019 - Dados iniciais Processados. . . . .	49
4.5	Distribuição agregada dos erros dos dados adicionais e com pré-processamento adicional. . . . .	52
4.6	Resultados das previsões de cimento para 2019 - Dados adicionais. . . . .	53
4.7	Explicabilidade <i>waterfall</i> para o estado do Pará. . . . .	55
4.8	Explicabilidade <i>waterfall</i> para o estado do Bahia. . . . .	55
4.9	Explicabilidade <i>waterfall</i> para o estado do Goiás. . . . .	56
4.10	Explicabilidade <i>waterfall</i> para o estado do São Paulo. . . . .	56
4.11	Explicabilidade <i>waterfall</i> para o estado do Paraná. . . . .	57

5.1	Exemplo de arquitetura autoregressiva . . . . .	61
A.1	Distribuição agregada dos erros dos dados de benchmark. . . . .	67
A.2	Resultados das previsões de jul/2017 a dez/2019 - Dados benchmark. . . . .	68

## Lista de tabelas

3.1	Tabela de informações dos dados iniciais . . . . .	31
3.2	Tabela de informação de métodos de pré-processamento dos dados iniciais	32
3.3	Tabela de informações dos dados do INMET . . . . .	33
3.4	Tabela de informações dos dados socioeconômicos adicionais . . . . .	33
3.5	Tabela de informação de métodos de pré-processamento dos dados adicionais	34
3.6	Tabela tipologia e uso de hiperparâmetros . . . . .	37
3.7	Descrição das Tecnologias Utilizadas . . . . .	42
4.1	Resultados estatísticos do experimento com dados brutos. . . . .	44
4.2	Resultados estatísticos do experimento com dados iniciais com pré- processamento adicional. . . . .	47
4.3	Resultados estatísticos do experimento com dados adicionais e com pré- processamento adicional. . . . .	51
5.1	Tabela de resultados totais. . . . .	59
A.1	Resultados estatísticos do experimento com dados brutos. . . . .	66



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Estudo de Caso: Indústria de Cimento - Demanda . . . . .	2
	Objetivo desse trabalho . . . . .	2
<b>2</b>	<b>Conceitualização</b>	<b>5</b>
2.1	Séries temporais . . . . .	5
2.1.1	Características inerentes . . . . .	5
2.1.2	Construções de séries temporais . . . . .	6
2.1.3	Extrapolação como predição . . . . .	6
2.2	Pré-processamento de dados . . . . .	7
2.2.1	Substituição por último valor observado, valores de fora do domínio ou moda . . . . .	7
2.2.2	Interpolação linear . . . . .	8
2.2.3	Interpolação Polinomial . . . . .	9
2.2.4	Transformações . . . . .	10
2.2.5	Extrapolação por regressão linear . . . . .	11
2.2.6	Recuo de entradas e média móvel . . . . .	12
2.2.7	Imputação com sazonalidade . . . . .	13
2.3	Aprendizado de máquina . . . . .	14
2.3.1	Definição e arcabouço para o aprendizado de máquina . . . . .	14
2.3.2	Aprendizado vs. Design . . . . .	15
2.4	Redes neurais de aprendizado de máquina . . . . .	15
2.4.1	Rede Neural Feed Forward - FFNN . . . . .	15
2.4.2	Rede Neural Recorrente - RNN . . . . .	18
2.5	Problemas comuns de treinamento de modelos . . . . .	20
2.5.1	Sobreajuste . . . . .	20
2.5.2	Dados insuficientes . . . . .	21
2.5.3	Dados imprecisos . . . . .	21

2.6	Boas práticas de treinamento de modelos de aprendizado de máquina . . .	22
2.6.1	Entendendo o domínio do problema . . . . .	22
2.6.2	Escolha de dados de entrada . . . . .	22
2.6.3	Ajustes de hiperparâmetros . . . . .	23
2.6.4	Métodos de regularização . . . . .	23
2.6.5	Validação . . . . .	25
2.7	Avaliação do trabalho anterior . . . . .	25
2.8	Explicabilidade . . . . .	26
2.8.1	Interpretando Modelos Preditivos . . . . .	26
2.8.2	Valores SHAP . . . . .	27
<b>3</b>	<b>Metodologia</b>	<b>29</b>
3.1	Entendimento e decisões baseadas na dinâmica do consumo de cimento .	29
3.1.1	Divisão estadual . . . . .	29
3.1.2	Recuo dos dados de entrada . . . . .	30
3.2	Preparação dos dados e organização . . . . .	30
3.2.1	Dados e fontes . . . . .	31
3.2.2	Escolha de granularidade e janela de ação . . . . .	34
3.3	Pré processamento do alvo de predição . . . . .	35
3.4	Criação de coluna de entrada de sazonalidade média cíclica. . . . .	36
3.5	Estrutura e ajustes de hiperparâmetros . . . . .	36
3.6	Metodologia de treinamento e análise de resultados . . . . .	38
3.6.1	Pré treinamento . . . . .	38
3.6.2	Treinamento . . . . .	39
3.6.3	Validação . . . . .	40
3.7	Tecnologias utilizadas . . . . .	41
<b>4</b>	<b>Experimentos e Resultados</b>	<b>43</b>
4.1	Treinamento de modelos de predição . . . . .	43
4.1.1	Fase 1: Treinamento sobre dados iniciais sem pré processamento .	44
4.1.2	Fase 2: Treinamentos sobre dados iniciais com imputação de inter- polação e média móvel . . . . .	47
4.1.3	Fase 3: Treinamentos com dados socioeconômico adicionais e dados meteorológicos . . . . .	50
4.2	Explicabilidade usando métodos SHAP . . . . .	54
<b>5</b>	<b>Interpretação dos resultados</b>	<b>59</b>
5.1	Possíveis próximos passos para futuros trabalhos . . . . .	60

<b>6 Conclusão final</b>	<b>63</b>
--------------------------	-----------

## **Apêndices**

<b>A Caso referência - Benchmark</b>	<b>65</b>
--------------------------------------	-----------

<b>Referências</b>	<b>69</b>
--------------------	-----------



# Capítulo 1

## Introdução

Nas últimas décadas, os modelos baseados em aprendizado de máquina transcenderam as fronteiras da esfera acadêmica, tornando-se uma parte integral dos estudos de pesquisa e desenvolvimento em praticamente todos os setores da sociedade. Seu crescimento expressivo, tanto como campo de estudo quanto em aplicabilidade, deve-se ao fato de que esses modelos ultrapassaram as expectativas comuns de um modelo matemático estatístico em termos de capacidade de previsão, classificação e geração.

A capacidade superior dos modelos baseados em aprendizado, especialmente os neurais, de capturar relações complexas entre os dados de entrada e representar bem as características do domínio ao qual o dado alvo que queremos prever pertence tornou-se, em muitos casos, a primeira opção quando se trata de previsibilidade em domínios complexos cujos dados influenciadores são multivariados. Sua estrutura versátil aceita dados de entradas multidimensionais, ao contrário de modelos mais simples e univariados. Além disso, modelos mais sofisticados e profundos<sup>1</sup> como os das redes neurais recorrentes, são capazes de mimetizar uma espécie de "memória", permitindo que o modelo capture variações sutis em séries temporais.

Sua aplicação se intensifica em diversas áreas, algumas delas como: a medicina, em que modelos treinados com grande quantidade de dados são capazes de gerar diagnósticos precisos, receitar medicamentos bem determinados com doses mais eficazes (RAJKOMAR *et al.*, 2019); na área de previsão climática, onde o fornecimento de indicadores futuros dos processos naturais levam a incentivos de projetos de mitigações de emissão de poluentes ou adaptações para condições inevitáveis inerentes às dinâmicas do planeta (CHATFIELD, 2000) e, finalmente, em áreas econômicas, em que é inevitável a análise de qualquer elemento por um prisma multivariado de parâmetros.

Todavia, uma simplificação desse uso acelerado é muito bem delimitada por MULLAINATHAN e SPIESS (2017) que enxerga o sucesso do aprendizado de máquina como uma mudança de paradigma na busca por soluções de problemas. Iniciamos de um ponto de partida procedural, buscando deduzir todas as regras que compõem o problema por meio de encadeamentos lógicos, para uma abordagem empírica, utilizando métodos indutivos para extrair relações de dados de forma automática; ou seja, os dados revelam suas informações

---

<sup>1</sup> Denominados, atualmente, como modelos de *Deep Learning* (Aprendizado profundo)

constituintes e relacionais.

## 1.1 Estudo de Caso: Indústria de Cimento - Demanda

O cimento é um material de uso tão intenso que pode ser considerado um dos pilares da sociedade moderna. Sendo o ingrediente chave na produção de concreto, está presente em todos os esforços construtivos relacionados à infraestrutura e moradia da civilização humana. Seu consumo mundial em 2020 foi equivalente a aproximadamente 4,16 bilhões de toneladas (SNIC, 2020) sendo, então, um dos materiais mais consumidos pela sociedade global. Sua importância é tão significativa que pode, por si só, servir como indicador de crescimento econômico de um país.

É de grande interesse para a indústria de cimento aumentar sua eficácia no atendimento da demanda, considerando que o cimento é um material com um prazo de validade relativamente curto, noventa dias, quando armazenado nas condições necessárias (NBR 16697, 2018). Após esse período, é possível reensaiá-lo, mas, caso não atenda às especificações necessárias, deve ser descartado. Isso impõe à indústria fornecedora a necessidade de um planejamento estratégico eficaz. Tanto a demanda quanto o posicionamento de suas plantas industriais e locais de armazenamento precisam ser cuidadosamente estudados.

Outro ponto, que torna crítica a necessidade de maior eficácia no planejamento é o fato dessa grande produção gerar uma das principais contribuições de emissões antropogênicas de CO<sub>2</sub> e desperdício de resíduos sólidos, especialmente na ausência de políticas eficientes de coprocessamento<sup>2</sup> (ARAUJO, 2020).

O aprimoramento da eficácia na produção e distribuição do cimento é de vital importância, não apenas para a indústria fabricante, mas também para toda a sociedade. À medida que as indústrias aperfeiçoam suas cadeias de produção, melhorando sua logística e reduzindo o desperdício, temos, não somente benefícios da redução de custos, mas também contribuição para a diminuição da poluição. Isto seria um importante avanço para uma indústria que atende um mercado com tendência de crescimento constante (IGHALO e ADENIYI, 2020).

### Objetivo desse trabalho

O objetivo deste trabalho é aprimorar modelos neurais recorrentes, como as redes Long-Short Term Memory (LSTM), no contexto da previsão de série temporal de consumo de cimento por unidade federativa brasileira e a gerar resultados de predição assertivos para o consumo referente ao intervalo de um ano. Isso será realizado através do uso de métodos de imputação e processamento de dados, boas práticas de treinamento e regularização de modelos de aprendizado de máquina, bem como pela incorporação de novos dados relevantes ao contexto do problema.

---

<sup>2</sup> O coprocessamento é uma técnica na qual os resíduos são empregados para substituir matérias-primas e/ou combustíveis na indústria de cimento.

Pretende-se avaliar o impacto no desempenho do modelo ao utilizar essas práticas em diferentes fases. Inicialmente, as fases serão conduzidas sem métodos sofisticados de imputação de dados faltantes e pré-processamento por média móvel, sendo esses métodos posteriormente introduzidos ao longo do processo para avaliação de seu impacto geral.

Não é objetivo, desse trabalho, uma análise do estado da arte do uso de aprendizado de máquina, comparação entre diversas arquiteturas de modelos preditivos ou avaliação dos resultados produzidos por combinações de diferentes tipos configurações de hiperparâmetros. Já foi tomado como partida o que a literatura atual considera como os modelos de melhor performance, dentre os neurais, para previsão de séries temporais, ou seja, modelos de rede neural recorrente como LSTM, stacked-LSTM e Bidirectional-LSTM (SANGIORGIO e DERCOLE (2020); BARRERA-ANIMAS *et al.* (2022)).

É, também, objetivo deste trabalho, buscar um melhor entendimento do domínio do problema através de recurso investigativo denominado *explicabilidade* dos modelos de aprendizado de máquina utilizando avaliação por valores SHAP (SHapley Additive exPlanations) que procuram avaliar a contribuição de cada variável de entrada no resultado final produzido pelo modelo.

É devido mencionar que, salvo as exceções referenciadas, todas as figuras apresentadas nesta monografia são de autoria do autor.



# Capítulo 2

## Conceitualização

### 2.1 Séries temporais

Quando lidamos com a previsão de eventos ou manifestações que ocorrem no futuro ao longo de uma escala de tempo, estamos lidando com extrapolação de série temporal.

#### 2.1.1 Características inerentes

Séries temporais são dados coletados através de observações sequenciais em um espaço de tempo com intervalos pré-definidos e organizados de forma cronológica. Independentemente se a variável observada for de natureza contínua ou discreta, a essência da série temporal demanda algum processo de discretização.

Podemos selecionar três características como atributos principais das séries temporais:

- **Tendências:** Pode ser definida como a identificação de um comportamento de variação que ocorre quando analisada uma janela de eventos de tamanho considerável. Em uma análise unidimensional, como o consumo de cimento, esse comportamento, quando identificado, pode ser de queda, crescimento ou estagnação. Podendo ser, o comportamento de queda ou crescimento, linear, exponencial ou logarítmico.
- **Sazonalidade:** Apesar do contexto mais amplo da palavra estar ligado a estações do globo terrestre, podemos estender esse conceito para uma série de padrões que são repetidos, nos valores das séries temporais, em intervalos cíclicos. Esse padrão não necessita ser uniforme, contudo pode ser identificado por agregação no determinado ciclo, tendo seus padrões e variâncias evidenciados.
- **Ruído:** São flutuações irregulares cuja ocorrência não pode ser contextualizada pela tendência e sazonalidade. Dessa forma, dentro do conjunto de dados, elas podem ser descartadas; no entanto, sempre há a possibilidade de que sejam parte integrante de um padrão não evidenciado ou de um conjunto de dados insuficiente para tal análise.

Autores como HAN *et al.* (2012) e o CHATFIELD (2000) separam o conceito de sazonalidade de "*movimentos cíclicos*" ou "*outras variações cíclicas*" formalizando, deste modo, um quarto atributo e isso é etimologicamente correto. Contudo, é evidentemente que a primeira pertence a um subconjunto de todos os tipos de variações cíclicas<sup>1</sup>.

### 2.1.2 Construções de séries temporais

Sendo dados coletados através de observações sequenciais em um espaço de tempo com intervalos pré-definidos, independentemente se a variável observada for de natureza contínua ou discreta, a essência da série temporal demanda algum processo de discretização. Esse processamento pode ser feito, de acordo com CHATFIELD (2000) dos seguintes modos:

- i. Sendo suas amostras coletadas em intervalos de tempo arbitrários para o caso de variáveis de natureza contínua.
- ii. Por agregação de dados coletados de forma irregular em um intervalo de tempo regular. Como, por exemplo, informações de consumo ou venda de um produto por mês.
- iii. Ou quando a variável observada já é, naturalmente, discreta e eventual. Por exemplo, pagamentos mensais parciais realizados sobre um montante a base de juros.

A grande maioria das variáveis observadas e delimitadas por uma série temporal apresentam um elevado grau de dependência entre passado e futuro, portanto, sua análise, fundamentalmente, deve levar em conta sua ordenação temporal.

### 2.1.3 Extrapolação como predição

Prever o futuro, seja com uma sofisticada rede neural ou uma simples presunção que o último valor observado se repetirá, é uma forma de extrapolação de dados. Uma extrapolação gera, seguindo algum tipo de lógica, uma continuidade de informações na série temporal em um período cronológico não observado. Isso é válido tanto para extrapolar o futuro quanto o passado.

Independente de qual modelo seja usado sobre o dado presente, a extrapolação sempre assume uma hipótese: a de que o conjunto de padrões atuais da séries, independente dos mesmos terem sido identificados ou não, se repetirão nos dados extrapolados, isto é, que as nuances intrínsecas dos dados observados e suas relações não mudaram durante o período cujo dado foi extrapolado. Essa particularidade sempre coloca a predição de dados em um território instável. Contudo, por termos o conhecimento dessa inevitabilidade, podemos também organizar o planejamento em torno dessa variabilidade considerando os erros de medidas e suas médias estatísticas.

A análise das séries temporais, segundo CHATFIELD (2000), tem quatro objetivos primordiais:

---

<sup>1</sup> Dado o caráter fortemente sazonal do presente trabalho, esses termos serão tratados como intercambiáveis ao decorrer do mesmo.

1. **Descrição:** Descrever os dados pelos seus atributos estatísticos ou visualmente através de gráficos.
2. **Modelagem:** Procurar uma representação matemática para o fenômeno observado que mimetize o seu comportamento esperado através de representações numéricas. Quando não conseguimos exprimir esse modelo deterministicamente por uma fórmula convencional temos, um caso muito adequado para o uso do recurso de aprendizado de máquina como ferramenta principal (ABU-MOSTAFA *et al.*, 2012).
3. **Previsão:** Com um modelo que represente satisfatoriamente o comportamento esperado do domínio estudado, podemos partir, não somente para a previsão de valores futuros na sequência temporal, como também, a análise das variáveis que constituem a base problema.
4. **Controle:** Com os resultados gerados pelo modelo e conhecimento dos elementos que determinam o resultado, torna-se possível adequar tomadas de decisões de acordo.<sup>2</sup>

## 2.2 Pré-processamento de dados

Os dados são a matéria prima para previsão de valores futuros em séries temporais. Todavia, para alcançar resultados precisos, é fundamental analisá-los, organizá-los e adaptá-los ao propósito específico. O esforço de trabalho nesse sentido é denominado pré-processamento de dados.

Um dos casos mais comuns e essenciais de pré-processamento é o de imputação de dados em sequências em que os mesmos estão ausentes. Modelos neurais, apesar de sua versatilidade em lidar com diferentes dimensões de dados de entrada, enfrentam dificuldades ao lidar com variações dimensionais durante o treinamento contínuo.

### 2.2.1 Substituição por último valor observado, valores de fora do domínio ou moda

Substituição é um dos tipos mais triviais de imputação de dados. E a mais simples substituição é quando repetimos o último valor de dado observado ou o próximo valor observado. Essas substituições apresentam custo computacional ínfimo, e são indicados a casos com poucos valores faltantes ou onde as questões temporais, de tendência e sazonalidade, não foram identificadas ou não se formalizam na granularidade dos dados em questão.

Em um ambiente mais classificatório, onde os valores de dados não apresentam forte correlação com o passado, pode ser efetuada uma substituição por moda, ou valor mais comum dentro do ciclo ou estação, em um caso em que a sazonalidade é evidente.

---

<sup>2</sup> Para o caso do cimento, isso permite a indústria adequar suas cadeias de produção de modo a aumentar eficiência e diminuir desperdícios. Permite também um planejamento mais assertivos sobre estratégias de implementação de fábricas ou estratégias de vendas.

No aprendizado de máquina, nos casos de modelos mais sofisticados, as vezes, podemos nos dar ao luxo de confiar na capacidade do modelo captar essas entradas faltantes sejam elas pontuais ou de janelas de dados faltantes. Uma prática de imputação acrescenta um valor de fora do domínio (e.g.: -1) nos valores faltantes e é otimista em relação a capacidade do modelo reconhecer esses valores como ausentes de correlação e então ignorá-los nos pesos dos resultado de sua saída.

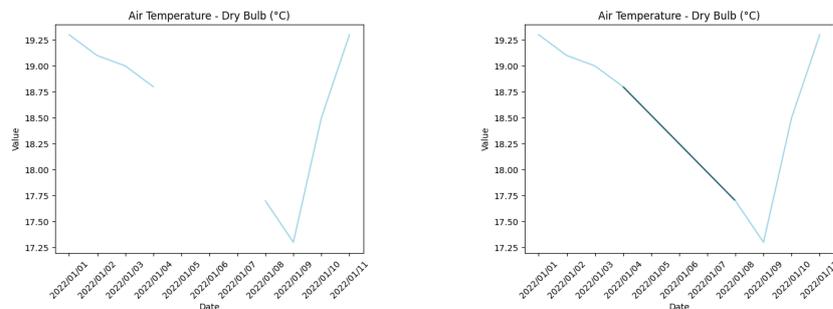
## 2.2.2 Interpolação linear

A interpolação linear consiste em imputar dados baseados em uma função linear entre dois pontos de dados observados em torno da sequência faltante e que estejam imediatamente antes e depois da mesma<sup>3</sup>. O Segmento de reta é delimitado segundo a fórmula:

$$y = y_{\text{anterior}} + (x - x_{\text{anterior}}) \cdot \left( \frac{y_{\text{posterior}} - y_{\text{anterior}}}{x_{\text{posterior}} - x_{\text{anterior}}} \right) \quad (2.1)$$

Onde:

- $x$  é o ponto dentro da sequência cronológica que queremos estimar,
- $y$  é o valor que queremos estimar,
- $x_{\text{anterior}}$  e  $x_{\text{posterior}}$  são as posições dos pontos imediatos observados na ordem cronológica,
- $y_{\text{anterior}}$  e  $y_{\text{posterior}}$  são os valores dos pontos imediatos, anterior e posterior.



(a) Série com dados faltantes.

(b) Dados imputados por interpolação linear.

**Figura 2.1:** Exemplo de interpolação linear.

A imputação por interpolação linear valoriza os dados observados ao passo que produz uma relação linear entre eles que age de forma similar a uma média. Esse método é eficiente computacionalmente e não gera ruídos com os dados novos. Contudo, a interpolação linear ignora completamente contextos mais amplos e próprios das séries temporais como as tendências anteriores e posteriores entre a janela de dados faltantes e as sazonalidades (2.1.1).

<sup>3</sup> Quanto somente queremos interpolar um único ponto, com seu antecessor e sucessor  $(x - 1)$  e  $(x + 1)$  presentes a interpolação linear age como média de ambos pontos imediatos.

### 2.2.3 Interpolação Polinomial

Outro tipo de interpolação por substituição de valores seguindo a regra de uma função é a interpolação polinomial. Esse tipo de imputação gera um polinômio:

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_dx^d \quad (2.2)$$

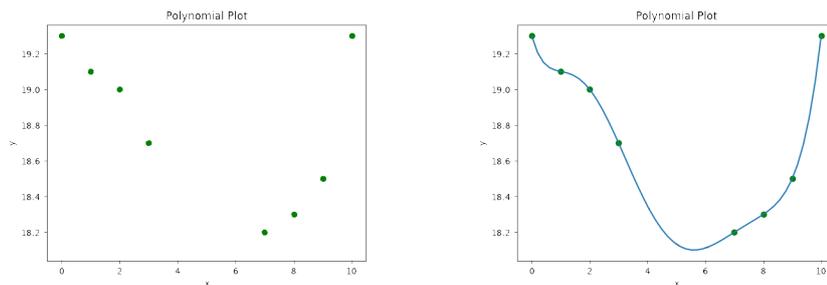
Utilizando parte dos pontos observados anteriores ao trecho faltante e parte dos pontos posteriores ao mesmo.

Existem vários métodos para interpolação polinomial, e um deles envolve o uso da matriz de *Vandermonde* para obter soluções numericamente estáveis com complexidade algorítmica de  $O(n^2)$ , em oposição à complexidade  $O(n^3)$  da eliminação gaussiana (WIKIPEDIA (2023)). Esta abordagem é apenas uma das muitas técnicas usadas no ajuste polinomial<sup>4</sup>.

O método de criação desses polinômios, usado neste trabalho, através da biblioteca *NumPy: Numerical Python* (2022), envolve o método dos mínimos quadrados. Em que procuramos reduzir o resultado da fórmula (2.3):

$$E = \sum_{i=1}^n [y_i - P(x_i)]^2 \quad (2.3)$$

- $n$  é o número de dados disponibilizados.
- $x_i$  e  $y_i$  são as coordenadas de cada um desses pontos no plano cartesiano.
- $P(x_i)$  é o valor previsto pelo polinômio a ser servido como equação base.
- $E$  é, portanto, a média dos erros quadrados de cada ponto.



(a) Série com dados faltantes.

(b) Interpolação polinomial criada.

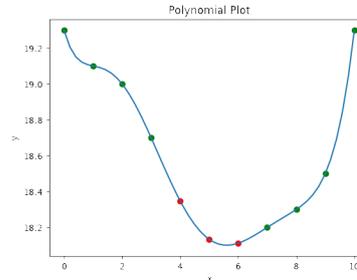
**Figura 2.2:** Exemplo de interpolação polinomial.

Isso gera os coeficientes de modo a diminuir o erro entre o polinômio e os pontos de dados observados oferecidos como base. Com esses coeficientes, geramos o polinômio da seguinte forma:

<sup>4</sup> Os métodos específicos empregados por bibliotecas como o `numpy.polyfit` do *NumPy: Numerical Python* (2022) para cálculo polinomial podem ser diferentes e muitas vezes são otimizados tanto para eficiência computacional quanto para estabilidade numérica.

$$P(x) = (x - c_1)(x^2 - c_2) \dots (x^n - c_n) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (2.4)$$

Com essa função em mãos, simplesmente imputamos os valores de  $P(x_f)$  para cada  $x_f$  faltante.



(a) Série com dados imputados.

Figura 2.3: Exemplo de interpolação polinomial com dados imputados.

A grande vantagem da interpolação polinomial frente à linear é que ela considera o contexto, ou melhor, as tendências posteriores e anteriores dos dados presentes. Deste modo, temos uma interpolação que leva em conta tendências; no entanto, ela ainda ignora completamente sazonalidades caso as mesmas não forem representadas por essas tendências preservadas.

## 2.2.4 Transformações

Transformações de dados estão mais ligadas a um pré-processamento com o intuito de melhorar o desempenho, ou seja, facilitar o treinamento modelo preditivo, a capturar as variações e relações entre os dados fornecidos. Além disso, auxilia na coordenação de atualizações em muitas camadas (GOODFELLOW *et al.*, 2017) durante os ajustes da operação de gradiente descendente.

A normalização, também conhecida como escala mínimo-máximo<sup>5</sup>, dimensiona uma série de dados para um intervalo específico, geralmente entre 0 e 1. Sua fórmula é:

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (2.5)$$

Onde:

- $X_{\text{norm}}$  é o valor normalizado.
- $X$  é o dado original da sequência selecionada.
- $\min(X)$  é o menor valor encontrado nessa sequência.
- $\max(X)$  é o maior valor encontrado nessa sequência.

<sup>5</sup> Min-max scaling

A *normalização* mantém a relação entre os dados intacta, colocando todas variáveis em uma mesma escala preservando as proporções. No entanto, um grande inconveniente do uso da normalização é a distorção causada por ruídos (outliers) presentes nos dados. Essa distorção pode prejudicar a captura de relações pelo modelo sobre a série de dados.

A *padronização*, por sua vez, também conhecida como escalonamento z-score, realiza uma transformação nos dados com o objetivo de tornar o conjunto resultante em um de média 0 e desvio padrão 1. Isso é alcançado com a seguinte operação:

$$Z = \frac{X - \mu}{\sigma} \quad (2.6)$$

Sendo:

- $Z$  é o z-score.
- $X$  é o dado original da sequência selecionada.
- $\mu$  é a média dos valores da sequência.
- $\sigma$  é o desvio padrão do conjunto de valores da sequência.

A padronização é menos afetada por outliers em comparação a normalização e, portanto, é mais indicada para séries de dados mais ruidosas. No entanto, a relação entre os dados é distorcida, o que pode resultar em perda de interpretabilidade da sequência. A escolha entre essas transformações deve levar em consideração a natureza dos dados e suas relações. Em muitos casos, a escolha é feita de forma empírica, observando e selecionando a transformação que resulta no melhor desempenho preditivo do modelo (QIAO *et al.*, 2019).

### 2.2.5 Extrapolação por regressão linear

Vimos, anteriormente, que a previsão do futuro é um tipo de extrapolação. No entanto, em certas ocasiões, precisamos imputar dados relacionados ao passado. Uma decisão tomada para o treinamento é o intervalo temporal dos dados a serem fornecidos ao modelo para aprendizado. No entanto, nem todos os dados desse conjunto podem cobrir o intervalo requerido. Por exemplo, podemos decidir que o intervalo de treinamento será de dados referentes aos meses de 2003 até 2019. No entanto, um dos tipos de dados escolhidos, para compor o conjunto de treinamento, tem informações disponíveis somente no intervalo entre 2005 e 2019. Dessa forma, podemos decidir entre imputar esses dados com valores indicativos de ausência (2.2.1) ou podemos extrapolá-los usando métodos como o de regressão linear.

Cada situação pode requerer uma abordagem diferente e uma simples regressão linear é, portanto, um método simples de extrapolação que pode ser eficaz para um curto período e possui um custo computacional relativamente baixo.

$$y = b_0 + b_1x \quad (2.7)$$

Sendo:

- $y$  é a variável de determinado ponto que desejamos imputar.
- $x$  é o ponto da série temporal onde a previsão deve ocorrer.
- $b_0$  é o valor da variável onde a reta intercepta o eixo  $y$ .
- $b_1$  é a inclinação (slope) da equação linear.

A chave da regressão está em determinar  $b_0$  e  $b_1$  de forma a obter uma reta cujo erro quadrado em relação aos pontos seja o menor possível. A fórmula para o cálculo da inclinação é:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.8)$$

Onde:

- $n$  é o número de dados presentes.
- $x_i$  e  $y_i$  são as coordenadas de cada um desses pontos no plano cartesiano.
- $\bar{x}$  e  $\bar{y}$  são as médias de posição na série e seu valor, respectivamente, de cada dado presente no conjunto.

O conjunto de dados usado para gerar a regressão não necessariamente precisa ser o conjunto total disponível. É possível escolher uma janela menor mais próxima ao intervalo a ser extrapolado para priorizar, desse modo, as relações entre os dados mais imediatos a esse intervalo.

## 2.2.6 Recuo de entradas e média móvel

Em muitos casos, desejamos reduzir o ruído na sequência de dados das séries temporais usada como entrada para o modelo de previsão. No entanto, também queremos preservar as variações dos dados de modo a não perder informações cíclicas e outras pequenas nuances. Um recurso interessante para isso é a média móvel. A média móvel<sup>6</sup> substitui os dados por uma média de valores pré-estabelecida que naturalmente inclui o próprio valor.

O tipo mais comum de média móvel é a média móvel simples (SMA - Simple Moving Average). Ela leva em consideração, para cada dado, o próprio valor e uma janela anterior. O valor da média é imputado na posição, para cada posição, como a média de todos esses dados:

$$SMA(t) = \frac{X(t) + X(t-1) + X(t-2) + \dots + X(t-n+1)}{n} \quad (2.9)$$

Sendo:

- $SMA(t)$  é o valor da média móvel simples para a posição  $t$ .

<sup>6</sup> Muitos autores, como [PYLE \(1999\)](#) e o [CHATFIELD \(2000\)](#), também citam o uso de média móvel como uma simples, porém eficaz, ferramenta de análise de tendências em séries temporais.

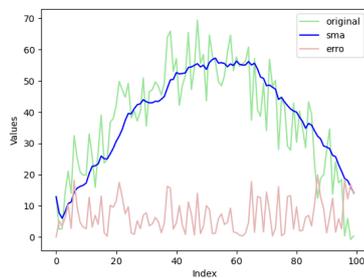
- $X(t)$  é o valor real para o dado na posição  $t$  na sequência.
- $n$  é o número de pontos delimitados pela janela de dados escolhida.

A realização da média móvel por SMA causa um efeito colateral denominado recuo de entradas (*lagged features*). Esse recuo às vezes é desejável quando um tipo de entrada tem uma correlação maior com o dado alvo quando está atrasada. Quando esse atraso não é desejado, talvez, por outros procedimentos de recuo já terem sido realizados, podemos considerar o uso de uma média móvel centrada. Sua única diferença em relação à simples é que ela leva em consideração o passado e o futuro na mesma proporção, ou seja, o valor alvo encontra-se no meio da janela de sequência escolhida. Ou seja, para um  $n$  ímpar, temos:

$$CMA(t) = \frac{X(t - \lfloor \frac{n}{2} \rfloor) + X(t - \lfloor \frac{n}{2} \rfloor + 1) + \dots + X(t) + \dots + X(t + \lfloor \frac{n}{2} \rfloor - 1) + X(t + \lfloor \frac{n}{2} \rfloor)}{n} \quad (2.10)$$

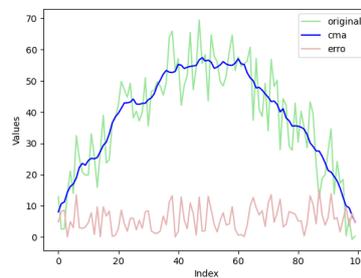
Onde  $CMA(t)$  é o valor da média móvel centrada para a posição  $t$ .

Essa operação tende a perder seu efeito à medida que o intervalo escolhido se torna indisponível perto do limite atual da série temporal. Fazendo-se necessário, para o algoritmo que esteja implementando uma CMA, reduzir a janela conforme ela vai se tornando indisponível ou, simplesmente não realiza-la para os pontos nessas condições.



(a) Média móvel comum (SMA<sup>a</sup>).

<sup>a</sup> Simple Moving Average



(b) Média móvel centrada (CMA<sup>a</sup>).

<sup>a</sup> Centered Moving Average

Figura 2.4: Exemplo de médias móveis.

### 2.2.7 Imputação com sazonalidade

Como já vimos, características inerentes a séries temporais são as tendências, sazonalidades e ruídos. Ao identificarmos esses fatores podemos subdividi-los no que é denominado de decomposição de séries temporais (CHATFIELD, 2000). Os dados de séries temporais são modelados, simplificadaamente, pela relação:

$$X_t = \mu_t \cdot i_t \cdot \varepsilon_t \quad (2.11)$$

Aqui estão as componentes deste modelo:

- $\mu_t$  representa a tendência ou componente de longo prazo.
- $i_t$  é a componente de sazonalidade.
- $\varepsilon_t$  denota a perturbação aleatória ou residual no tempo  $t$ .

Para facilitar a análise e modelagem, este modelo multiplicativo pode ser transformado em uma forma totalmente aditiva usando uma transformação logarítmica:

$$\log(X_t) = \log(\mu_t) + \log(i_t) + \log(\varepsilon_t) \quad (2.12)$$

Existem diversas formas de algoritmos que buscam separar e delimitar essas contribuições. Contudo, identificado um padrão cíclico, podemos manipulá-lo, salientando ou em alguns casos até descartando-os, nas chamadas “*séries dessazonalizadas*”.

## 2.3 Aprendizado de máquina

### 2.3.1 Definição e arcabouço para o aprendizado de máquina

Qualquer domínio cujo comportamento pode ser representado por um modelo de aprendizado de máquina (ABU-MOSTAFA *et al.*, 2012) deve possuir um espaço  $X$ , denominado espaço de entrada, um conjunto  $D$  de pares entrada-saída no formato  $(x_1, y_1), \dots, (x_n, y_n)$ , e uma função desconhecida  $f$  no formato  $f : X \rightarrow Y$ . Modelos de aprendizado de máquina são aqueles que, com técnicas recursivas de treinamento, geram um modelo  $g : X \rightarrow Y$  que pretende se aproximar de  $f$ . O conjunto de todos os modelos, independentemente de quão bem se aproximam de  $f$ , é denominado conjunto de Hipóteses  $H$ .

Em relação ao conjunto  $D$ , quando temos conhecimento de todos os pares  $(x_i, y_i)$  no conjunto de treino, ou seja, para cada  $x_i$ , conhecemos os valores de  $y_i$ , estamos em uma situação de aprendizado supervisionado. Nesse caso, o modelo é treinado tendo disponível para seu algoritmo de treinamento os alvos  $y_i$  correspondentes a cada entrada  $x_i$ .

No caso de um modelo de previsão de séries temporais, como neste trabalho, temos uma situação de aprendizado supervisionado para o treinamento. A previsão de valores futuros, por outro lado, não se encaixa nessa classificação, pois desejamos usar  $x_j$  antes de termos os  $y_j$  correspondentes, em uma situação externa ao treinamento. Denominamos essa aplicação, de previsão de valores futuros com base em entradas do presente e do passado, de “teste fora da amostra” ou “teste no conjunto externo”, cujos erros compõem o conjunto  $E_{out}$ .

Para buscar o melhor modelo no conjunto  $H$ , uma estratégia é treinarmos os modelos utilizando amostras conhecidas, mas não utilizados pelo algoritmo de treinamento para ajuste de parâmetros, os erros dessa estimativas são chamados erros dentro da amostra  $E_{in}$ . O modelo que apresentar o melhor desempenho, de acordo com os resultados de  $E_{in}$ , é então escolhido para estimar  $E_{out}$ .

Independentemente dos resultados de treinamento em  $E_{in}$ , o modelo só demonstra sua qualidade se tiver um bom desempenho em  $E_{out}$ . Esse é o objetivo fundamental de praticamente todas as aplicações de aprendizado de máquina em situações de previsão de

valores futuros de séries temporais. A capacidade de um modelo baseado em aprendizado de máquina de performar em um conjunto de dados externo ao treinamento é referida como capacidade de *generalização* do modelo.

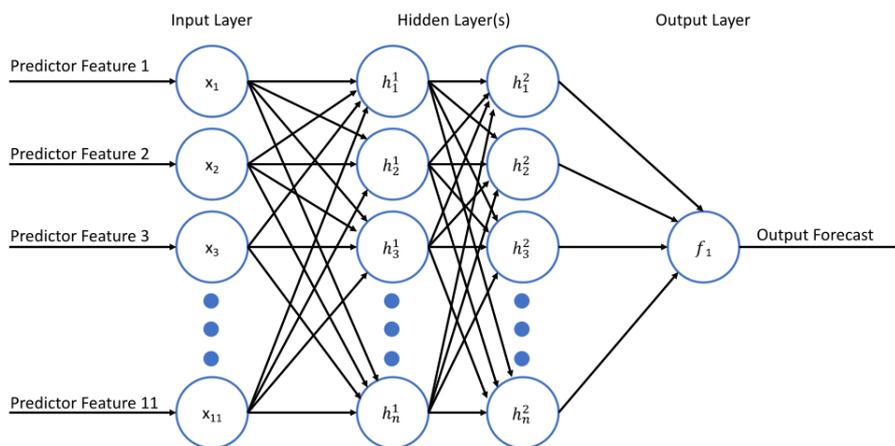
### 2.3.2 Aprendizado vs. Design

Uma distinção pertinente explicada por [ABU-MOSTAFA \*et al.\* \(2012\)](#), é a diferença de aprendizado e design. Quando consideramos regras de aplicação pré-determinadas, baseadas em particularidades, factuais ou não, do domínio estudado e a aplicamos no processo preditivo, seja na totalidade ou em alguma parcela de um fluxo lógico, temos então um design. Este design, ainda pode ser considerado parte do que hoje definimos como inteligência artificial<sup>7</sup>. O aprendizado extrai suas relações dos dados oferecidos automaticamente, enquanto as regras de aplicação de design são delimitadas com algum conhecimento prévio.

## 2.4 Redes neurais de aprendizado de máquina

### 2.4.1 Rede Neural Feed Forward - FFNN

Uma rede neural artificial, ou perceptron de multicamada, como o próprio nome sugere, possui uma rede conectada com mais de uma camada de perceptrons interconectados. As entradas são recebidas por uma camada de entrada que transmite o sinal de cada entrada para cada perceptron da camada seguinte, até chegar na camada final de saída, responsável por devolver o resultado da computação. Denominamos as camadas entre as camadas de entrada e saída de camadas escondidas.



**Figura 2.5:** Rede neural feed-forward<sup>a</sup>. Figura de [BARRERA-ANIMAS \*et al.\* \(2022\)](#)<sup>8</sup>

<sup>a</sup> Predictor Feature n (Tipo de dado de entrada); Input Layer (Camada de entrada); Hidden Layer (Camadas ocultas) e Output Layer (Camada de saída)

<sup>7</sup> O aprendizado de máquina é amplamente considerado um subconjunto do que é conhecido como inteligência artificial (I.A.) ([GOODFELLOW \*et al.\*, 2017](#))

O nome "feed-forward" é aplicado a essa estrutura devido ao fato de que cada camada posterior se alimenta da saída da camada anterior em um fluxo, que apesar de ser ramificado, é direcionado. Esse aumento de complexidade permite ao modelo captar um maior número de relações entre os dados de entrada. Isso ocorre devido ao valor do dado de entrada ter uma trajetória muito mais variada, contribuindo de diferentes formas em diferentes fluxos de informação numérica.

$$a_i = \sigma(W_i \cdot a_{i-1} + b_i) \quad (2.13)$$

Onde:

- $a_i$  é o vetor de saída na camada  $i$ .
- $W_i$  é a matriz de pesos associada à camada  $i$ .
- $a_{i-1}$  é o vetor de ativação da camada atual, ou saída da camada anterior (camada  $i - 1$ ).
- $b_i$  é o vetor de viés da camada  $i$ .
- $\sigma$  é a função de ativação.

O aprendizado automático ocorre através das alterações dos pesos da rede através da operação de *gradiente descendente*, que pode ser definido como um algoritmo que procura minimizar medidas de erro através de iterações contínuas que navegam pela superfície do espaço de erro em direção a um mínimo<sup>9</sup>. Uma função de erro que pode ser usada como base é a do erro quadrado:

$$L_i = e_i^2 = (y_i - \hat{y}_i)^2 \quad (2.14)$$

Onde:

- $L_i$  é o valor da função de perda para a predição  $i$ .
- $e_i$  é o erro da predição  $i$ .
- $y_i$  é o valor alvo ou valor observado.
- $\hat{y}_i$  é o valor previsto ou resultado do modelo.

Talvez, a forma mais comum de gradiente descendente seja o estocástico<sup>10</sup>. Importante não confundir com os otimizadores usados em bibliotecas de treinamento, muitos otimizadores, como o Adam, utilizam gradiente descendente estocástico combinando-o

<sup>8</sup> Licenciada sob CC BY 4.0. <https://s100.copyright.com/AppDispatchServlet?publisherName=ELS&contentID=S266682702100102X&orderBeanReset=true>

<sup>9</sup> É comum dizermos que o gradiente descendente "navega" em direção aos menores valores da superfície de erro, contudo o gradiente aponta para a maior inclinação de aumento de função do ponto selecionado. O algoritmo leva em consideração isso para seguir em direção oposta.

<sup>10</sup> Stochastic Gradient Descent - SGD

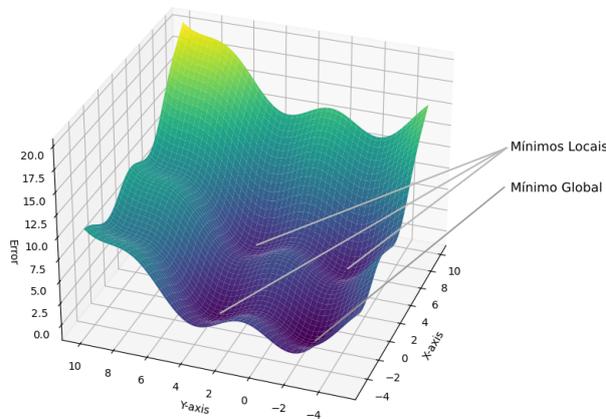
com diversas outras medidas. Esse gradiente descendente atualiza um parâmetro da rede neural na forma da regra:

$$w_{\text{novo}} = w_{\text{antigo}} - \eta \cdot \frac{\partial L}{\partial w} \quad (2.15)$$

Onde:

- $w_{\text{novo}}$  é o novo peso atualizado.
- $w_{\text{antigo}}$  é o peso antigo usado na atualização.
- $\eta$  é a taxa de aprendizagem, que corresponde à proporção da atualização do peso em cada iteração.
- $\frac{\partial L}{\partial w}$  é o gradiente, que é computado usando a regra da cadeia sobre a função de perda (*loss function*).

O cálculo de redução do erro pode ser representado por uma superfície (para o caso multivariado) de erro onde há vales, cumes, pontos de sela e platôs. Dentre esses valores, alguns têm seus pontos de menor valor representando mínimos locais. Esses mínimos locais são extremamente indesejáveis para o processo de redução de erro, pois nos desviam do potencial de melhora que se configura como mínimo global, isto é, o menor valor de erro da função de erro utilizada.



**Figura 2.6:** Superfície da função de custo e exemplo de pontos de mínimo local e global.

O comportamento do SGD<sup>11</sup> é iterativo e envolve atualizar os valores até ser alcançada uma convergência. O nome "estocástico" é atribuído devido não ser considerado todo conjunto de parâmetros em cada iteração. Esse procedimento, além de acelerar o processo, origina ruído na atualização de parâmetros que pode ser favorável para variabilidade do processo, diminuindo as chances do valor convergir para um mínimo local.

<sup>11</sup> Stochastic Gradient Descent

## 2.4.2 Rede Neural Recorrente - RNN

Um tipo sofisticado de Rede Neural são as redes neurais recorrentes (RNN). Sua principal diferenciação se dá por meio de uma estrutura de fluxo de dados que promove a criação de uma espécie de memória. Esses fluxos funcionam como loops que registram valores cujos pesos são sensíveis ao contexto ou sequência dos dados de entrada.

Uma formalização simplificada de uma RNN pode ser definida da seguinte forma:

$$H_t = f(H_{t-1}, X_t) \quad (2.16)$$

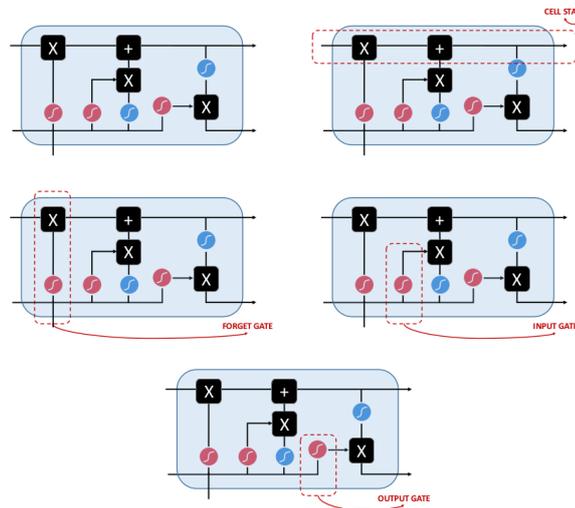
$$\hat{Y}_{t+h|t} = g(H_t) \quad (2.17)$$

Onde:

- $H_t$  é o estado denominado oculto (*hidden state*).
- $f$  e  $g$  são funções próprias a estrutura da rede recorrente.
- $\hat{Y}_{t+h|t}$  é a predição referente ao valor  $Y_{t+h}$ , levando em consideração o último estado  $t$ .

Essa formulação  $\hat{Y}_{t+h|t}$  é a ideia chave da dependência sequencial que as redes recorrentes capturam.

A principal variante desse tipo de rede é denominada Rede LSTM (Long Short-Term Memory). Na LSTM, o fluxo de recorrência é controlado dinamicamente pela união das células LSTM e é sensível à ordem das entradas.



**Figura 2.7:** Delimitações da Rede LSTM. Figura de MASINI *et al.* (2023)<sup>12</sup>

Suas ramificações de fluxo são:

<sup>12</sup> Licenciada sob CC BY 4.0. <https://arxiv.org/abs/2012.12802>

- *Input gate (Porta de entrada)*: Responsável por decidir quais informações serão adicionadas ao *cell state (estado da célula)* atual  $C_t$  utilizando o estado anterior  $H_{t-1}$  e a nova entrada  $X_t$  através de uma função de ativação.
- *Forget gate (Porta de esquecimento)*: Responsável por delimitar quais informações do estado anterior  $C_{t-1}$  e input atual  $H_t$  serão "esquecidas" e quais continuarão no fluxo de "memória".
- *Output gate (Porta de saída)*: Delimita quais informações serão passadas para o *hidden state (estado oculto)* e quais seguirão para o próximo estado de célula  $C_t$ .

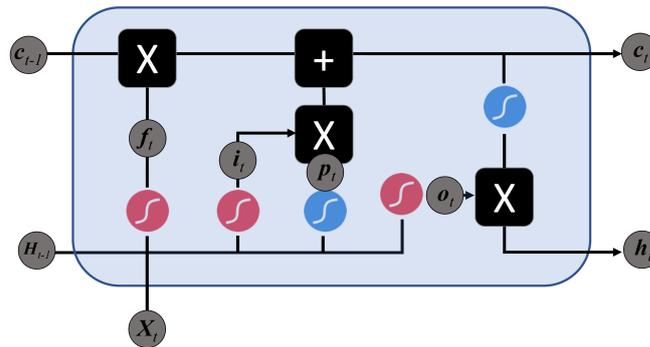


Figura 2.8: Célula LSTM. Figura de MASINI et al. (2023)<sup>13</sup>

A figura (2.8), representa uma célula de rede LSTM. Elas geralmente são concatenadas uma com as outras. Sua sequência de execução pode ser exemplificada pela série de operações abaixo:

$$f_t = \sigma(W_f X_t + U_f H_{t-1} + b_f) \quad (2.18)$$

$$i_t = \sigma(W_i X_t + U_i H_{t-1} + b_i) \quad (2.19)$$

$$o_t = \sigma(W_o X_t + U_o H_{t-1} + b_o) \quad (2.20)$$

Tanto a porta de entrada  $i_t$ , esquecimento  $f_t$  e saída  $o_t$  recebem tanto o vetor de entrada atual  $X_t$  quanto o vetor correspondente ao estado oculto anterior,  $H_{t-1}$ , e os multiplicam pela suas respectivas matrizes de pesos. Somando-os todos junto aos seus respectivos *bias* e processando-os pela função de ativação  $\sigma$ .

$$p_t = \text{Tanh}(W_c X_t + U_c H_{t-1} + b_c) \quad (2.21)$$

A etapa  $p_t$  age de forma semelhante, contudo processa o resultado por uma função de ativação de tangente hiperbólica.

<sup>13</sup> Licenciada sob CC BY 4.0. <https://arxiv.org/abs/2012.12802>

$$c_t = (f_t \odot c_{t-1}) + (i_t \odot p_t) \quad (2.22)$$

O estado da célula a ser atualizado  $c_t$ , por sua vez, é formado pelo produto termo a termo<sup>14</sup> da saída da porta de esquecimento  $f_t$  e o estado da célula anterior,  $c_{t-1}$  e os somam, usando a mesma operação de produto, feita sobre a porta de entrada  $i_t$  e o estado intermediário  $p_t$ .

$$h_t = o_t \odot \text{Tanh}(c_t) \quad (2.23)$$

O estado oculto atual, a ser usado pela próxima célula LSTM da sequência, recebe a tangente hiperbólica do estado da célula atualizado  $c_t$ , calculado na mesma etapa  $t$ , e realiza um produto termo a termo com a o resultado do cálculo atual da porta de saída  $o_t$ .

As equações acima são apresentadas, sendo:

- $\sigma$ : função de ativação sigmóide.
- $\text{Tanh}$ : função de ativação tangente hiperbólica.
- $W_f, W_i, W_o, W_c, U_f, U_i, U_o, U_c$ : são os pesos,  $W$  são os referentes a entrada atual e  $U$ , a serem ajustados pelo gradiente descendente.
- $b_f, b_i, b_o, b_c$ : são os *bias* correspondentes à cada porta ou estado.
- $H_{t-1}$ : É o valor de entrada referente ao estado oculto antigo  $t - 1$ .
- $H_t$  É os valores de entrada referentes ao estado atual  $t$ .
- $i_t, f_t, p_t, c_t, o_t, h_t$  representam o *Input gate (Porta de entrada)*, *forget gate (Porta de esquecimento)*, *cell state (estado da célula)*, *updated cell state (estado da célula atualizado)*, *Output gate* e *Hidden State (estado oculto)*, respectivamente.

## 2.5 Problemas comuns de treinamento de modelos

### 2.5.1 Sobreajuste

Segundo [ABU-MOSTAFA et al. \(2012\)](#), sobreajuste (*overfitting*) é um fenômeno que ocorre quando encontramos um resultado adequado ou uma explicação para um problema que representa muito bem um cenário conhecido. Contudo, esse modelo perde a validade quando exposto a dados inéditos. Em outras palavras, o sobreajuste ocorre quando um modelo é escolhido no conjunto de hipóteses  $H$  devido à sua grande performance ou, melhor dizendo, baixo valor de erro em amostra de treinamento ( $E_{in}$ ). No entanto, quando submetido a uma série externa de dados inéditos, temos uma baixa performance ou erro fora da amostra ( $E_{out}$ ) alto. O problema é que muitas vezes recorremos ao aprendizado de máquina quando o problema é muito complexo para o uso de técnicas mais triviais, e o mesmo autor afirma que existe um aumento probabilístico de sobreajuste quando há um aumento na complexidade do alvo.

<sup>14</sup>  $\odot$  Produto *Hadamard*.

Portanto, os três principais fatores que favorecem o sobreajuste são:

- Complexidade do alvo: um alvo complexo e com muitas nuances favorece o sobreajuste. Por exemplo, um polinômio de alto grau é difícil de ser encontrado com poucos dados, aumentando a probabilidade de seus pontos coincidirem com os de outro polinômio.
- Conjunto reduzido de dados de treinamento: Um conjunto reduzido oferece pouca informação necessária para o modelo evitar o sobreajuste, ao passo que também é mais fácil "encaixar" um modelo errado em uma série de dados pequena.
- Ruído nos dados de entrada: Entradas incorretas de dados também contribuem para o sobreajuste devido à dificuldade do modelo em reconhecer esses *outliers*.

As redes neurais são muito eficientes em encontrar uma hipótese adequada para o conjunto de dados fornecidos para treinamento. Deste modo, o sobreajuste torna-se um dos principais problemas na busca de um modelo que tenha um bom desempenho em uma amostra inédita. Muitos dos outros problemas que veremos na sequência são definidos como tal devido propiciarem a ocorrência de sobreajuste, ou seja, há uma relação intrínseca entre esse último e a qualidade e quantidade dos dados (ABU-MOSTAFA *et al.*, 2012).

## 2.5.2 Dados insuficientes

Quando dispomos de poucos dados para treinar nossos modelos, enfrentamos sérios riscos de desenvolver um modelo hipotético suscetível ao sobreajuste. Isso ocorre porque não há amostras suficientes para que o modelo seja capaz de capturar as verdadeiras relações entre os dados do domínio do problema. O modelo treinado nessas condições tenderá a gerar relações erradas ou correlações que, com a entrada de dados novos, acabem mostrando-se deturpadas. Por exemplo, um tipo de dado de entrada pode apresentar uma forte correlação com o dado alvo no trecho temporal delimitado para treinamento, entretanto, com avanço das observações na série cronológica, esse dado pode revelar-se não correlacionado no contexto real de causa e efeito. O modelo treinado sob essa limitação é propenso a capturar essa correlação inicial e tornar-se excessivamente dependente, em termos de valores de resultados, dessa entrada enganosa.

## 2.5.3 Dados imprecisos

Dados são a matéria-prima de todo modelo de aprendizado de máquina supervisionado. O modelo de aprendizado ajusta seus parâmetros de modo a melhor representar a relação  $(X_1, y_1), \dots, (X_n, y_n)$  de entradas e saídas. Portanto, a quantidade, bem como a qualidade desses dados de entrada, é essencial para o treinamento do modelo. Contudo, nem sempre há certeza dos dados de entrada e a relação que o modelo gera entre eles refletem bem o comportamento do domínio que modelo busca reproduzir. Uma variável de dado de entrada com forte correlação com o dado de saída pode contribuir para que o modelo tenha uma boa performance de treino. No entanto, somente em sua extensa aplicação em amostras externas, ela será julgada como fator determinante de causa-consequência em relação à variável alvo. Dados incorretos introduzem ruídos no modelo durante o treinamento, comprometendo sua performance ao distorcer os parâmetros.

## 2.6 Boas práticas de treinamento de modelos de aprendizado de máquina

### 2.6.1 Entendendo o domínio do problema

O primeiro passo em qualquer empreendimento de treinamento de modelos de aprendizado de máquina é adquirir conhecimento sobre o domínio em que o alvo a ser previsto está inserido. Esse esforço é fundamental por diversos aspectos. A escolha de dados relevantes dentro do domínio estudado aumenta as chances do modelo captar relações interessantes entre eles, aproximando o resultado previsto da realidade (AGGARWAL, 2018).

Ter um entendimento aprofundado do problema facilita a escolha das técnicas mais apropriadas para processar os dados. Por exemplo, em certos tipos de dados, a presença de ruído pode ser relevante, especialmente quando a variância é pequena. Em contraponto, também com conhecimento prévio, podemos perceber que variáveis ruidosas podem interferir negativamente nos resultados.

Outro caso relevante, que é muito usado especialmente no mercado financeiro (KROLLNER *et al.*, 2010), ocorre quando há conhecimento prévio de que uma variável candidata a parâmetro de treinamento tem seu maior efeito, no resultado preditivo, quando a mesma é atrasada (2.2.6) em algum intervalo de tempo determinado. Isso se deve à natureza cronológica das séries temporais e seus efeitos no domínio. Por exemplo, uma variável correspondente a investimentos em educação pode ser atrasada em alguns anos em um cenário relacionada a previsão de avanços tecnológicos ou consumo de livros.

### 2.6.2 Escolha de dados de entrada

A escolha dos tipos de dados a serem utilizados, como mencionado anteriormente, depende do conhecimento sobre o domínio do problema. No entanto, existem outros aspectos relacionados à escolha de dados específicos que devem ser cuidadosamente considerados para evitar problemas que possam resultar em distorções na predição do modelo.

Em primeiro lugar, os dados devem ser provenientes de fontes confiáveis, com seus processos de obtenção, processamento e publicação bem documentados. As distorções causadas por dados incorretos podem ser difíceis de mensurar devido à dificuldade em diagnosticar a causa do problema (2.5.3), e os dados incorretos podem não ser tão evidentes quanto dados ausentes.

Outra consideração relevante para a escolha dos dados está relacionada ao seu processamento. A granularidade dos intervalos de amostragem dos dados, no caso de séries temporais, geralmente envolve unidades de medida temporais, como horas, dias, semanas, meses e anos. É importante considerar a granularidade mais apropriada de acordo com o objetivo da predição. No entanto, uma granularidade mais fina sempre oferece mais oportunidades para modelos de aprendizado automático, pois eles se beneficiam do aumento da quantidade de dados fornecidos para treinamento<sup>15</sup>.

---

<sup>15</sup> AGGARWAL (2018) considera, como os dois fatores mais proeminentes para o avanço do uso de ML, nas últimas décadas, o aumento do poder computacional e o aumento ao acesso de dados.

### 2.6.3 Ajustes de hiperparâmetros

Podemos definir hiperparâmetros como toda a configuração que controle ou influencie o comportamento do algoritmo de aprendizado automático (GOODFELLOW *et al.*, 2017), mas que não sejam alterados pelo mesmo durante o processo de aprendizado.

Esse ajuste é essencial devido a redes neurais serem estruturas complexas que possuem configurações ajustáveis para diversos tipos de necessidades a fim de manter o caráter versátil das mesmas.

O ajuste de hiperparâmetros pode ser realizado de várias maneiras: manualmente, por tentativa e erro; de forma automatizada, utilizando técnicas como otimização baseada em métodos de busca, incluindo pesquisa em grade, busca aleatória e otimização bayesiana; ou ainda fundamentado na teoria de estudos prévios. Esses métodos, quando combinados, fazem parte do que é conhecido atualmente como AutoML<sup>16</sup>.

Um dos métodos mais simples e mais usados é o da busca em grade (*grid search*) (HUTTER *et al.*, 2019). Ele é um método que delimita os hiperparâmetros em listas de seus possíveis valores e gera resultados para cada uma das combinações.

De nada adianta um ajuste extensivo se for utilizado de forma ineficaz, o que pode resultar em um sistema com severos problemas de *overfitting*. Portanto, é importante definir um critério que determinará o desempenho dos parâmetros ajustados. Assim, o ideal para o modelo é que seu ajuste seja validado por um conjunto de dados não utilizado pelo algoritmo de aprendizado durante o ajuste automático de parâmetros internos.

"Os hiperparâmetros não devem ser ajustados usando o mesmos dados usados para gradiente descendente. Em vez disso, uma parte dos dados é apresentada como dados de validação, e o desempenho do modelo é testado no conjunto de validação com várias opções de hiperparâmetros." (AGGARWAL, 2018, p. 125)

### 2.6.4 Métodos de regularização

Métodos de regularização são procedimentos técnicos e sistemáticos que auxiliam modelos de aprendizado a evitar o sobreajuste, promovendo melhor desempenho de generalização. A maioria dos métodos envolve a modificação do comportamento do algoritmo de treinamento para penalizar certas condições ou impor restrições que o forcem a se comportar de maneira mais benéfica.

"Regularização é qualquer modificação que fazemos em um algoritmo de aprendizagem que visa reduzir seu erro de generalização, mas não seu erro de treinamento." (GOODFELLOW *et al.*, 2017)

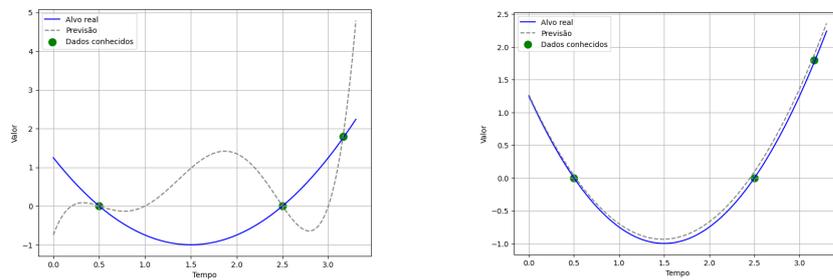
Dentre os procedimentos, destacam-se os que penalizam as normas dos parâmetros. Esses métodos adicionam uma penalidade de norma na forma de um parâmetro  $\Omega(\theta)$ . A fórmula geral é a seguinte:

<sup>16</sup> O estudo de ajustes de hiperparâmetros é amplo e, por si só, se formaliza como uma área de estudo em *machine learning*.

$$\tilde{J}(\theta; X, y) = J(\theta; X, y) + \alpha\Omega(\theta) \quad (2.24)$$

Onde:

- $\tilde{J}(\theta; X, y)$  é a função objetivo<sup>17</sup> regularizada com seus parâmetros ajustados.
- $J(\theta; X, y)$  é a função objetivo sem uso de regularização.
- $\alpha$ , com  $\alpha \in [0, \infty)$ , é um hiperparâmetro que atua como o coeficiente de contribuição do termo de penalidade  $\Omega$  na função objetivo.
- $\theta$  são os parâmetros do modelo.



(a) Previsão sobre o efeito do sobreajuste. (b) Previsão após métodos de regularização.

**Figura 2.9:** Exemplo de correção de sobreajuste por regularização<sup>a</sup>.

<sup>a</sup> O Gráfico a esquerda representa um polinômio desenvolvido que, apesar de adequar-se perfeitamente aos dados conhecidos, claramente não representa bem o alvo real. A figura a direita, por sua vez, aproxima muito melhor o resultado desejado, devido a regularização.

É importante ressaltar que esses procedimentos de penalização ocorrem na maioria dos casos sobre os parâmetros, ou pesos  $W$ , das redes neurais e não sobre seus bias  $b$ . Dois regularizadores de parâmetros que trabalham com penalidades são a regularização de parâmetros L2, também conhecida como decaimento de peso<sup>18</sup>, e a regularização de parâmetros L1, que tende a propiciar uma solução considerada mais esparsa (GOODFELLOW *et al.*, 2017), com alguns parâmetros tendendo a zero e, portanto, atua como uma selecionadora de entradas.

Outras formas de regularização impõem restrições aos modelos durante sua execução. Um exemplo é a técnica chamada *dropout*<sup>19</sup>, que aleatoriamente redefine o valor de um conjunto preestabelecido de variáveis para zero, entre uma camada intermediária e sua sucessora, promovendo diversificação do modelo, pois o modelo não pode contar com a mesma variável de entrada em cada iteração<sup>20</sup> de aprendizado.

<sup>17</sup> Objetivo por ser a função que o modelo procura otimizar.

<sup>18</sup> Também referida como regressão de crista ou regularização Tikhonov

<sup>19</sup> No *TensorFlow: An Open Source Machine Learning Framework for Everyone* (2023), esse método é introduzido por uma camada, também denominada Dropout, entre duas camadas normais da rede. [https://www.tensorflow.org/api\\_docs/python/tf/keras/layers/Dropout](https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dropout)

<sup>20</sup> Denominadas na área de *machine learning* como epochs (épocas)

### 2.6.5 Validação

Para termos um maior grau de certeza de que o modelo está performando bem em um ambiente de testes que não foi usado no treinamento, ou seja, está generalizando bem, podemos realizar esse tipo de teste durante o treinamento na chamada validação cruzada<sup>21</sup>. Nesse método, os dados são divididos em partes não sobrepostas, de modo que a cada iteração do treinamento, o modelo seja testado com uma parte dos dados separados para teste. Esse algoritmo favorece um monitoramento da performance do modelo, que é mais valioso do que um cálculo de perda sobre uma entrada que foi usada para o treinamento, evitando, portanto, o sobreajuste.

A base da validação consiste em inserir dados como entrada em um modelo treinado, ou em processo de treinamento, e medir o erro de sua saída frente ao valor real com alguma métrica. Isso para um caso de aprendizado supervisionado. Existem diversas formas de medirmos esse erro, algumas delas são:

- MAE (Mean Absolute Error): Calcula a média dos erros absolutos registrados em cada previsão. Sua fórmula é:

$$MAE = \frac{1}{n} \sum |Y - \hat{Y}|, \quad (2.25)$$

Onde  $n$  é o número de amostras,  $Y$  é o dado real e  $\hat{Y}$  é o dado previsto.

- MAPE (Mean Absolute Percentage Error): Similar ao MAE, contudo, seu erro medido em cada previsão é percentual. Pode ter seu valor final multiplicado por 100 para ser exibido como porcentagem ou como fração. Seu cálculo é:

$$MAPE = \frac{1}{n} \sum \left| \frac{(\hat{Y} - Y)}{\hat{Y}} \right| \times 100\%. \quad (2.26)$$

- RMSE (Root Mean Squared Error): É uma variante dos cálculos de erro que tende a penalizar maiores variações no cálculo da média de erros, uma vez que eleva ao quadrado a diferença do erro antes de aplicar uma raiz quadrada sobre sua média. RMSE oferece uma informação interessante para os casos em que grandes erros ou outliers não podem ser tolerados, e sua ocorrência acaba sendo diluída em um cálculo de MAE com muitas previsões. Ele pode ser obtido através do cálculo abaixo:

$$RMSE = \sqrt{\frac{1}{n} \sum (\hat{Y} - Y)^2}. \quad (2.27)$$

## 2.7 Avaliação do trabalho anterior

Não somente é natural, como é essencial na construção do conhecimento e desenvolvimento de trabalhos atuais, a análise crítica de trabalhos anteriores, tanto a nível de meros

<sup>21</sup> k-fold cross-validation

erros e acertos, quanto no aproveitamento de bases já iniciadas ou exploradas. Deste modo, é importante mencionar que ano passado houve um trabalho muito similar a este também sob a supervisão do Prof. Dr. Marcelo Finger, realizado pela aluna Julia Leite<sup>22</sup>.

O trabalho anterior, embora tenha focado na diversificação dos tipos de modelos preditivos e suas diferentes performances, formalizou um bom ponto de partida para o presente devido sua decisão de escolha dos tipos de dados de entrada, muito significativos dentro do contexto da complexa relação de consumo de cimento pela sociedade brasileira. Também forneceu um cenário já estabelecido para análise crítica de sua metodologia, escolhas de treinamento e validação que inspiraram esse trabalho a explorar mais a fundo boas práticas de validação e treino de modelos neurais.

A abrangência de sua abordagem<sup>23</sup> na questão arquitetural dos modelos preditivos, recursos estatísticos de avaliação de erro<sup>24</sup> e transformações de dados servem de ponto inicial de pesquisa para esforços que procurem entender o tema e explorá-lo.

Na modelagem de modelos preditivos temos que levar em conta, não somente aspectos técnicos da confecção do mesmo e as particularidades do domínio do alvo de predição, como também, a factibilidade e usabilidade do modelo gerado em um cenário real necessidade de predição. O trabalho anterior utilizou um recuo (2.2.6) de entrada de somente um mês para dados de granularidade mensal e um ano para dados de granularidade anual. Isso o colocou em território de difícil aplicação devido o atraso do tempo de liberação desses dados e inaplicabilidade na previsão de mais de um mês no futuro (dado que para a previsão do próximo mês  $m + 1$  seria necessário os dados atuais  $m$  e para previsão de  $m - 2$  os próprios dados de  $m + 1$ , ou seja, dados do futuro).

## 2.8 Explicabilidade

### 2.8.1 Interpretando Modelos Preditivos

Os modelos neurais estão entre os modelos preditivos de maior complexidade. Entender como eles funcionam em um cenário multivariado pode se tornar uma tarefa muito difícil devido à grande quantidade de permutações do tipo presente e não presente que teríamos que aplicar sobre suas variáveis de entrada, bem como à sua presença ou ausência. Os benefícios de um modelo interpretável são valiosos e diversificados, pois fornece indicadores sobre a importância de determinadas variáveis em relação a outras, auxiliando nos processos de decisões, tanto para os mantenedores do modelo quanto, principalmente, para aqueles que trabalham no domínio ao qual o modelo foi projetado.

---

<sup>22</sup> "Previsão de consumo de cimento nos estados do Brasil usando métodos de aprendizado automático" <https://linux.ime.usp.br/~jleite/mac0499/>

<sup>23</sup> O trabalho anterior focou tanto métodos simples de predição como regressão linear até redes sofisticadas como LSTM e Bidirecional-LSTM

<sup>24</sup> O trabalho anterior usou as mesmas medidas de avaliação de erro (MAE, MAPE e RSME)

## 2.8.2 Valores SHAP

Valores SHAP atuam como um conjunto de medidas que buscam avaliar a importância atribuída a cada dado de entrada.

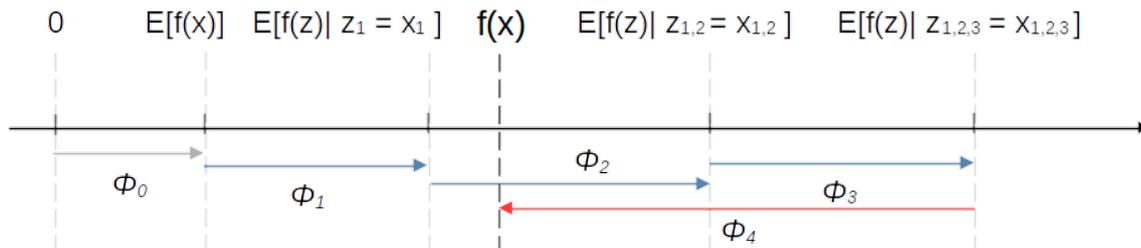
Para um determinado modo de previsão, o valor SHAP para uma entrada  $i$  é calculado da seguinte forma:

$$\phi(i) = \sum_{z' \subseteq N} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \setminus \{i\})] \quad (2.28)$$

Onde:

- $\phi(i)$  é o valor SHAP para a variável de entrada  $i$ .
- $f_x(z')$  é a previsão do modelo considerando um subconjunto de entradas  $z'$ .
- $f_x(z' \setminus \{i\})$  é a previsão do modelo para um subconjunto de entradas  $z'$  sem a entrada  $z'_i$  zerado ou anulada<sup>25</sup>.
- $z'$  representa um subconjunto do conjunto de tipos de entrada  $N$  não zeradas. No contexto dos valores SHAP,  $z'$  é um conjunto de recursos ou variáveis que está sendo considerado. A fórmula calcula o valor de Shapley para o  $i$ -ésimo recurso somando as diferenças sobre todos os subconjuntos possíveis  $z'$ .
- $M$  representa o número total de recursos no conjunto de tipos de entrada  $N$ <sup>26</sup>.

Não é objetivo deste trabalho aprofundar nas propriedades e teoremas que levaram a essa expressão. Entretanto, caso haja interesse, o trabalho de [LUNDBERG e LEE \(2017\)](#) as delimita detalhadamente. A figura (2.10) ilustra como esses valores se relacionam:



**Figura 2.10:** Figura explicativa de valores SHAP.

As variáveis  $\phi$  de 0 a 4 são os valores SHAP,  $f(x)$  é o resultado da previsão do modelo, e  $E[f(z)]$  é o valor-base, ou seja, o valor do resultado caso todas as variáveis estivessem presentes e com valores médios. Então  $E[f(z)|z_i = x_i]$  é o valor aproximado do modelo quando a variável  $z_i$  é definida como entrada com o valor  $x_i$ . Dessa forma, o valor SHAP  $\phi_i$

<sup>25</sup> Não foi colocado  $z'_i = 0$ , como no trabalho referência de [LUNDBERG e LEE \(2017\)](#), porque, como comentado em (4.2), bibliotecas como *SHAP (SHapley Additive exPlanations)* (2023) utilizam o conceito de "plano de fundo".

<sup>26</sup> A expressão  $\frac{|z'|!(M - |z'| - 1)!}{M!}$  é equivalente ao inverso de um coeficiente binomial  $\frac{1}{\binom{M}{|z'|}}$ .

é a diferença  $E[f(z)|z_i = x_i] - E[f(z)]$ . Essas esperanças são sensíveis à ordem em que os tipos de entrada são adicionados, e assim, em todas as ordenações possíveis, os valores SHAP surgem da média dos valores de  $\phi_i$  (LUNDBERG e LEE, 2017).

# Capítulo 3

## Metodologia

### 3.1 Entendimento e decisões baseadas na dinâmica do consumo de cimento

Como já foi introduzido, uma eficaz previsão de demanda de cimento por localidade é estrategicamente essencial para um planejamento eficaz da indústria de cimento. As decisões de projeto deste trabalho foram tomadas levando em consideração as condições da indústria de cimento, a necessidade de predição, a disponibilidade de dados públicos e os limites de tempo inerentes a um trabalho de pouco menos de um ano.

O consumo de cimento é um evento cuja pontualidade é incerta, deste modo o Sindicato Nacional da Indústria de Cimento (SNIC, 2020) reúne as informações disponibilizadas pelos seus participantes e utiliza agregação (2.1.2) para gerar uma série temporal de consumo de cimento por mês dividido por estado da federação brasileira.

Para a previsão, foram escolhidos os modelos de redes neurais do tipo LSTM. Sendo as do tipo perceptron multicamadas<sup>1</sup> somente utilizadas para a análise de explicabilidade por valores SHAP devido melhor compatibilidade.

Como as dinâmicas dos estados não são as mesmas, a previsão de consumo de cimento para cada estado contou com sua própria rede neural. Essa escolha foi tomada com o intuito de facilitar o modelo a captar relações próprias de cada estado, partindo do pressuposto de que determinada variável pode não apresentar o mesmo efeito no valor alvo em um modelo de previsão para um determinado estado federativo em comparação com outro<sup>2</sup>.

#### 3.1.1 Divisão estadual

Como vimos, a indústria carece de previsões precisas e bem definidas quanto ao seu espaço geográfico, sendo a sua setorização essencial para o planejamento de implantações

---

<sup>1</sup> Outro modo de referir-se a uma Rede Neural Feed-Forward

<sup>2</sup> Modelos complexos de redes neurais usados nesse trabalho são capazes de "aprender" essas nuances, contudo ao separarmos os casos manualmente, garantimos um esforço a menos, diminuindo a complexidade do problema

de plantas industriais e ajustes logísticos de transporte. Neste trabalho, foi escolhida a delimitação em unidades federativas da união, ou seja, estados. Isso se deve ao fato de que o SNIC já fornece dados de consumo de cimento por unidade da federação com granularidade mensal, como já mencionado, e o fato da disponibilidade da maioria dos dados públicos relacionados a índices socioeconômicos pertinentes à questão estudada apresentem seus valores divididos por estados<sup>3</sup>.

### 3.1.2 Recuo dos dados de entrada

Uma previsão de valores futuros não pode contar com dados do futuro para ser formulada. Precisamos prever o futuro com base nas informações do passado disponíveis no presente. Em muitos casos, dados de informações de caráter socioeconômicos são liberadas anualmente, mesmo que se configurem em granularidades menores (E.g.: trimestral, mensal etc.)<sup>4</sup>. Isso torna, o simples recuo de um ano impraticável, devido a diversas fontes, terem sua divulgação formalizada meses após o fim do ano em questão.

Portanto, todos os dados usados para a previsão são defasados em um intervalo de dois anos. Esse procedimento, além de ser necessário, também nos aproxima de um ambiente real de treinamento contínuo e previsão de intervalos de tempo anuais futuros. Ao estarmos no ano  $x$  temos informações de dados referentes ao ano  $x - 1$ , podendo, deste modo, iniciar a previsão de valores para o ano  $x + 1$ .

## 3.2 Preparação dos dados e organização

Este trabalho aproveitou da escolha de seleção de dados de entrada definidos no trabalho anterior,<sup>5</sup> Portanto, existem 15 colunas de dados de entrada relacionados a índices de desenvolvimento, índices socioeconômicos e valores quantitativos relacionados à população e produção de cimento<sup>6</sup>.

Além disso, foi planejado a adição de novas séries de dados de entrada para fins experimentais. O consumo de cimento é um fenômeno complexo, e muitas vezes não sabemos quem são seus fatores influentes e como eles afetam o resultado.

Conseqüentemente, foram adicionados dois tipos de séries de dados. Primeiramente, incluímos dados meteorológicos por estado. Isso foi feito por duas razões principais: devido a possibilidade das variações nas condições dos fenômenos atmosféricos podem influenciar o consumo de cimento, afetando sua durabilidade em situações em que os materiais não

---

<sup>3</sup> Alguns dados do IBGE podem ser encontrados em divisões de municípios, mas o processamento desses dados em escalas menores resultaria em um esforço desproporcional ao escopo da disciplina à qual este trabalho está inserido

<sup>4</sup> E.g.: O próprio consumo de cimento, que é o alvo de predição deste trabalho, é divulgado anualmente, apesar de ser organizado e agregado por ano

<sup>5</sup> "Previsão de consumo de cimento nos estados do Brasil usando métodos de aprendizado automático-  
<https://linux.ime.usp.br/~jleite/mac0499/>

<sup>6</sup> Para teste de referência de desempenho foram feitos testes iniciais sob moldes parecidos ao do trabalho anterior. E seus resultados podem ser conferidos no apêndice (A)

são armazenados adequadamente <sup>7</sup> ou porque foram vendidos no varejo e os compradores não seguiram as precauções de armazenamento adequadas. Em segundo lugar, espera-se que um modelo treinado com esses parâmetros se torne mais sensível às variações sazonais associadas aos fenômenos atmosféricos, aumentando a probabilidade do modelo capturar as variações sazonais no consumo de cimento em sí.

Outra série de dados adicionada refere-se a dados socioeconômicos relacionados a índices mais específicos, poder de compra e endividamento da população e riscos econômicos. Essa inclusão de dados econômicos adicionais baseia-se na análise dos relatórios anuais sobre o consumo e produção de cimento elaborados pelo SNIC. Atualmente, é amplamente reconhecido que o consumo de cimento está intrinsecamente ligado às condições econômicas da nação, até mesmo em nível de pessoa física, e é um fator determinante nesse contexto.

### 3.2.1 Dados e fontes

Abaixo, segue a tabela de dados iniciais:

Dado	Unidade	Divisão	Granularidade	Período Original
Consumo de cimento	mil ton.	Estadual	Mensal	2003-2022
Produção de cimento	mil ton.	Estadual	Mensal	2003-2022
Valor do Cimento <sup>a</sup>	R\$/Kg	Estadual	Mensal	1994-2022
Desemprego	%	Estadual	Trimestral	2002-2016 e 2014-2022
IDH	-	Estadual	Anual	1970-2021
PIB - Estadual	R\$ (mil.)	Estadual	Anual	1939-2020
PIB - Construção Civil	R\$ (mil.)	Estadual	Anual	1939-2020
PIB - Estadual per capita	R\$ (mil.) <sup>b</sup>	Estadual	Anual	1985-2022
PIB - Preços de Mercado	R\$ (mil.)	Estadual	Anual	1939-2020
NFSP	% PIB	Nacional	Mensal	1999-2022
ELCF	R\$ de 2010 (mi.)	Nacional	Anual	1947-2022
IGP-DI	%	Nacional	Mensal	1944-2022
INCC	%	Nacional	Mensal	1944-2022
IPCA	- <sup>c</sup>	Nacional	Mensal	1994-2022
População	Habitante	Nacional	Anual	1872-2022
SELIC	%	Nacional	Diário	1986-2022

**Tabela 3.1:** Tabela de informações dos dados iniciais

<sup>a</sup> Portland 32.

<sup>b</sup> Preços do ano 2010.

<sup>c</sup> Número-Índice (1993=100).

Esses dados iniciais são os mesmos utilizados no trabalho anterior. Optamos por reutilizá-los devido serem uma escolha adequada visto que representam um conjunto

<sup>7</sup> Por exemplo, quando um depósito não atende às especificações das normas brasileiras

tradicional de indicadores socioeconômicos de acesso público muito utilizados em diversas áreas de pesquisa.

Dado	Método <sup>a</sup> I	Método II	Fonte	URL
Consumo de cimento	IS	-	SNIC	I
Produção de cimento	-	-	SNIC	I
Valor do Cimento	IP	-	SNIC	I
Desemprego	IP	-	IBGE + BCB	II + III
IDH	IP	IL	IPEA	II
PIB - Estadual	-	-	IBGE	II
PIB - Construção Civil	-	-	IBGE	II
PIB - Estadual per capita	-	-	IBGE	II
PIB - Preços de Mercado	-	-	IBGE	II
NFSP	-	-	BCB	IV
ELCF	-	-	BCB	IV
IGP-DI	-	-	FGV	II
INCC	-	-	FGV	II
IPCA	IP	-	IBGE	V
População	IP	-	IBGE	II
SELIC	IL	-	IBGE	VI
IS	Imputação com Sazonalidade			
IL	Interpolação Linear			
IP	Interpolação Polinomial			
CMA	Média Móvel Centrada <sup>b</sup>			
I	<a href="http://www.cbicdados.com.br/">http://www.cbicdados.com.br/</a>			
II	<a href="http://www.ipeadata.gov.br/Default.aspx">http://www.ipeadata.gov.br/Default.aspx</a>			
III	<a href="https://www.bcb.gov.br/estatisticas">https://www.bcb.gov.br/estatisticas</a>			
IV	<a href="https://dadosabertos.bcb.gov.br/">https://dadosabertos.bcb.gov.br/</a>			
V	<a href="https://www.ibge.gov.br/estatisticas/">https://www.ibge.gov.br/estatisticas/</a>			
VI	<a href="https://www.debit.com.br/tabelas/">https://www.debit.com.br/tabelas/</a>			

**Tabela 3.2:** Tabela de informação de métodos de pré-processamento dos dados iniciais

<sup>a</sup> Método de pré-processamento usado para imputação sobre os dados de entrada

<sup>b</sup> Centered moving average - OBS.: Todos os dados, exceto o alvo consumo de cimento, foram processadas por média móvel centrada

Os dados meteorológicos foram todos retirados do INMET<sup>8</sup>, abaixo a relação:

Todos os dados meteorológicos foram retirados das bases públicas do INMET<sup>9</sup> e foram processadas por uma mistura de regressão linear, imputação por interpolação polinomial e, finalmente, média móvel centrada.

<sup>8</sup> Instituto Nacional de Meteorologia

<sup>9</sup> <https://portal.inmet.gov.br/>

Dado	Unidade	Divisão	Granularidade	Período Original
Precipitação	mm/dia	Estadual	Diário	(2001 2008)-2022
Pressão atmosférica	mB	Estadual	Diário	(2001 2008)-2022
Radiação global	kJ/m <sup>2</sup>	Estadual	Diário	(2001 2008)-2022
Temperatura do ar	°C	Estadual	Diário	(2001 2008)-2022
Temp. ponto de orvalho	°C	Estadual	Diário	(2001 2008)-2022
Umidade relativa do ar	%	Estadual	Diário	(2001 2008)-2022
Precipitação	m/s	Estadual	Diário	(2001 2008)-2022

**Tabela 3.3:** Tabela de informações dos dados do INMET

Os dados socioeconômicos adicionais, ou dados expandidos<sup>10</sup>, foram os da tabela seguinte:

Dado	Unidade	Divisão	Granularidade	Período Original
Custo m <sup>2</sup>	R\$	Estadual	Mensal	1999-2022
Depósito Poupança	R\$	Estadual	Anual	2003-2022
IDH - Educação	-	Estadual	Anual	1970-2021
IDH - Longevidade	-	Estadual	Anual	1970-2021
IDH - Renda	-	Estadual	Anual	1970-2021
NFSP - Fluxo Mensal	R\$ (milhões)	Nacional	Mensal	1999-2022
Operações de crédito <sup>a</sup>	R\$ (milhões)	Nacional	Mensal	1990-2022
EMBI Risco-Brasil	-	Nacional	Diário	1994-2022
FGTS	R\$ (mil)	Nacional	Mensal	1988-2022
Operações de crédito <sup>b</sup>	- <sup>c</sup>	Nacional	Mensal	2010-2022
PIB <sup>d</sup>	R\$ (mil)	Nacional	Mensal	1996-2022
PPC <sup>e</sup>	R\$ (mil)	Nacional	Anual	1940-2022

**Tabela 3.4:** Tabela de informações dos dados socioeconômicos adicionais

<sup>a</sup> Saldo da carteira de crédito

<sup>b</sup> Inadimplência da carteira de crédito

<sup>c</sup> Número-Índice (1995=100)

<sup>d</sup> Atividades imobiliárias

<sup>e</sup> Salário mínimo - paridade do poder de compra (PPC)

<sup>10</sup> Usamos, neste trabalho, a palavra expandido no sentido de que já foram incluídos anteriormente importantes dados socioeconômicos, contudo essa adição representa uma expansão de suas informações.

Dado	Método I	Método II	Fonte	URL
Custo m <sup>2</sup>	-	-	SIDRA	VII
Depósito Poupança	IL	-	BCB	II
IDH - Educação	IP	IL	IPEA	II
IDH - Longevidade	IP	IL	IPEA	II
IDH - Renda	IP	IL	IPEA	II
NFSP - Fluxo Mensal	-	-	BCB	IV
Operações de crédito	-	-	IBGE	II
EMBI Risco-Brasil	-	-	J.P. Morgan <sup>a</sup>	II
FGTS	-	-	IBGE	II
Operações de crédito	IP	IL	BCB	II
PIB <sup>b</sup>	-	-	IBGE	II
PPC <sup>c</sup>	-	-	IPEA	II
IL	Interpolação Linear			
IP	Interpolação Polinomial			
CMA	Média Móvel Centrada <sup>d</sup>			
II	<a href="http://www.ipeadata.gov.br/Default.aspx">http://www.ipeadata.gov.br/Default.aspx</a>			
IV	<a href="https://dadosabertos.bcb.gov.br/">https://dadosabertos.bcb.gov.br/</a>			
VII	<a href="https://sidra.ibge.gov.br/">https://sidra.ibge.gov.br/</a>			

**Tabela 3.5:** Tabela de informação de métodos de pré-processamento dos dados adicionais

<sup>a</sup> Banco de investimento multinacional americano e empresa de serviços financeiros - <https://www.jpmorgan.com.br/pt/about-us>

<sup>b</sup> Atividades imobiliárias

<sup>c</sup> Salário mínimo - paridade do poder de compra (PPC)

<sup>d</sup> Centered moving average - OBS.: Todos as entradas foram processadas por média móvel centrada

### 3.2.2 Escolha de granularidade e janela de ação

A escolha da granularidade decorreu da granularidade mínima oferecida pela fonte de dados de consumo de cimento, ou seja, o SNIC. Desta forma, todas as séries temporais dos dados de entrada usados para a predição foram postos dessa forma ou processados por interpolação para isto.

A janela de ação, ou janela escolhida para treinamento e predição, foi a correspondente ao início do ano de 2003 até o final do ano de 2019. Essa escolha se estabeleceu devido a maior presença de dados confiáveis durante esse período. O período das últimas duas décadas foi mais estável economicamente, pelo menos quando comparado ao referente a décadas anteriores que foram marcados por elevada inflação e adoção de políticas econômicas expressivas como a do plano real.

A escolha de data inicial de 2003, é devido a dois fatores: ausência de certos vetores de dados, principalmente os meteorológicos, em datas anteriores e proximidade da data em que o SNIC começou a disponibilizar seus relatórios anuais com informações mensais de

consumo de cimento<sup>11</sup>.

A data alvo de predição escolhida foi a correspondentes aos meses referente ao ano de 2019<sup>12</sup>

### 3.3 Pré processamento do alvo de predição

A coluna de dados de consumo de cimento é o nosso alvo de predição. Portanto, é essencial não alterá-la. No entanto, durante os anos de 2014, 2015 e 2016, não foi encontrado informações detalhadas na granularidade mensal. A única informação disponível é o consumo total anual.

Para lidar com essa falta de dados mensais, foi desenvolvido um procedimento de imputação que mantém os atributos essenciais da série temporal, preservando a precisão do consumo anual. Como mencionado anteriormente, a sazonalidade é um atributo crucial das séries temporais. Dada a informação anual de consumo, o objetivo foi reconstruir uma distribuição mensal que fosse consistente com o comportamento sazonal dos dados observados anteriormente.

O procedimento seguiu as etapas a seguir:

Primeiramente, calculamos a média dos valores de consumo de cimento para o conjunto de dados, agrupados por mês. Isso nos forneceu uma espécie de sazonalidade média, considerando que o consumo de cimento, geralmente, segue um ciclo anual. O cálculo foi o seguinte:

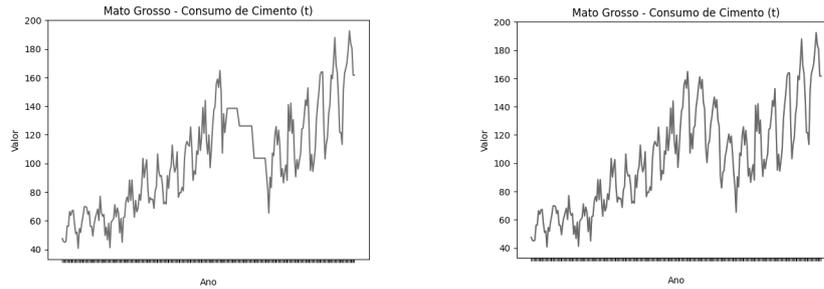
$$C_{\text{input}}(m, a) = C_{\text{anual}}(a) \times \frac{M_{\text{sazonal}}(m)}{\sum_{i=0}^{11} M_{\text{sazonal}}(i)} \quad (3.1)$$

Onde:

- $C_{\text{input}}(m, a)$  é o valor a ser imputado para o mês referente ao consumo do ano observado.
- $C_{\text{anual}}(a)$  é o consumo de cimento referente ao ano "a" disponível.
- $M_{\text{sazonal}}(m)$  é a média de consumo de cimento relativa ao mês "m" calculada dentro do intervalo observado.
- $\sum_{i=0}^{11} M_{\text{sazonal}}(i)$  é a soma de todas as médias mensais para cada um dos doze meses "i".

<sup>11</sup> O primeiro relatório é o correspondente ao ano 2001 e pode ser encontrados, juntos com os demais, em <http://snic.org.br/numeros-relatorio-anual.php>

<sup>12</sup> Descartamos 2020 e 2021 devido à incidência da pandemia do COVID-19 distorcer os dados de praticamente todos os setores da humanidade (NICOLA *et al.*, 2020) e 2022 por ausência de diversos dados devido à proximidade com a data deste trabalho (2023).



(a) *Série temporal com dados imputados por média.* (b) *Dados imputados com sazonalidade média.*

**Figura 3.1:** *Exemplo de imputação com sazonalidade média.*

O gráfico (a) mostra a imputação por média para o consumo de cimento do estado do Mato Grosso havendo ausência dos dados na granularidade de mês entre os anos de 2014 e 2016. A média do consumo anual é dividida pelos meses criando platôs nas séries. Em contrapartida, a imputação (b) leva em conta a sazonalidade média cíclica e, devido sua fórmula, mantêm o valor anual intacto.

### 3.4 Criação de coluna de entrada de sazonalidade média cíclica.

Nada impede a criação de tipos de entradas personalizados em machine learning. Somente deve, haver o cuidado de não utilizar informações que, teoricamente, o modelo ainda não teria acesso em um estudo de caso de teste como esse trabalho acadêmico. Devido a análise de que o modelo não está capturando muito bem as nuances de sazonalidade das séries registradas, além de introduzir dados meteorológicos, foi criado uma coluna de sazonalidade média cíclica com os dados de consumo de cimentos anteriores ao ano que temos por objetivo prever. Portanto essa, coluna foi criada agregando por mês a média do consumo de cimento daquele mês e é cíclica, pois foi repetido esse valor para cada conjunto de dados de entrada de treinamento correspondente a um ano.

$$SMC(m) = M_{\text{sazonal de 2003 à 2017}}(m) \quad (3.2)$$

Sendo:

- $SMC(m)$  o valor imputado para a coluna “Sazonalidade” ao mês  $m$ .
- $M_{\text{sazonal de 2003 à 2017}}(m)$  a média do consumo de cimento do mês  $m$  durante o período de 2003 à 2017.

### 3.5 Estrutura e ajustes de hiperparâmetros

Como estrutura dos modelos de predição, seguimos com baterias de testes utilizando rede LSTM, em específico a denominada stacked-LSTM que consiste em mais de uma

camada de células LSTM<sup>13</sup>, e é muito eficaz em capturar relações complexas (SANGIORGIO e DERCOLE (2020), CHUNG *et al.* (2014) e YU *et al.* (2019)).

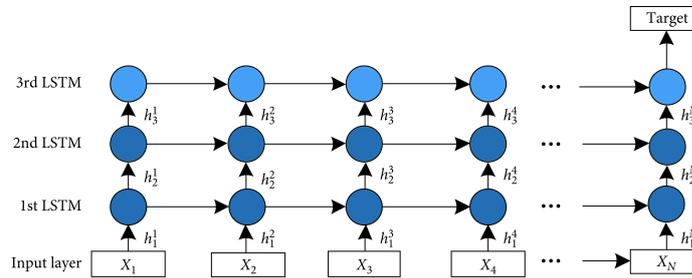


Figura 3.2: Stacked-LSTM. Figura adaptada de YU *et al.* (2019)<sup>14</sup>

O ajustes de hiperparâmetros foi feito de forma empírica sempre sobre resultados gerados sobre o corpo de amostra interna usado apenas para validação durante o treinamento do modelo e não ajustes de seus parâmetros internos ajustado pelo gradiente descendente (2.6.3). Parte de seus ajustes foram feitos *manualmente* por tentativa e erro, sendo reservado para o número de unidades (células) e número de camadas a busca em grade.

Hiperparâmetros	Valores	Ajuste
Camadas	2 e 3	Automatizado ( <i>grid search</i> )
Unidades	36, 72 e 144	Automatizado ( <i>grid search</i> )
Paciência ( <i>patience</i> )	10, 100 e 500	Manual
Descarte ( <i>dropout</i> )	0.1, 0.2 e 0.5	Manual
Maturação ( <i>start from epoch</i> )	0.1, 0.2 e 0.5	Manual
Taxa de Aprendizagem	0.001 e 0.005	Manual e Padrão ( <i>Default</i> )
Ativação ( <i>activation</i> )	ReLU e Tanh	Manual
Otimizador	Adam e SDG ( <i>experimental</i> )	Manual
Regularizador L2 ( $\alpha$ )	0.01	Padrão ( <i>Default</i> )
Épocas ( <i>epochs</i> )	1000, 5000, 10000	Manual

Tabela 3.6: Tabela tipologia e uso de hiperparâmetros

Foi descartada a possibilidade de automatizar todos os parâmetros da tabela (3.6) pelos seguintes motivos, além do elevado número relativo a permutação de todos valores de hiperparâmetros: o fato de que este trabalho optou por realizar um modelo preditivo para cada estado da união, ou seja, 26 estados mais distrito federal e pela natureza do otimizador estocástico que requer uma seleção de seus melhores resultados dentre uma série de tentativas.<sup>15</sup>

<sup>13</sup> Durante testes iniciais, para reconhecimento e familiarização com as ferramentas, foi decidido descartar a utilização de somente uma camada de células LSTM devido alta incidência de resultados fora do domínio (negativos)

<sup>14</sup> Licenciada sob CC BY 4.0. <https://www.hindawi.com/journals/sv/2019/2756284/#copyright>

<sup>15</sup> O treinamento de modelos necessários para uma automatização de busca de grade para a tabela (3.6) foi calculado em 3.499.200 treinamentos (25.920 permutações x 27 unidades federativas x 5 treinamentos de validação.)

Além disso o ajustes manuais por tentativa e erro ou com permanência de valores padrão (*Default*) não são, de nenhuma forma, uma opção inapropriada. *WEERTS et al. (2020)* apresentaram experimentos indicativos que a permanência de determinados hiperparâmetros em seu valor padrão não produz resultados inferiores e, até mesmo, podem produzir melhores resultados quando comparados a procedimentos iterativos de número limitado.

## 3.6 Metodologia de treinamento e análise de resultados

Ao longo do tempo compreendido por este trabalho foram feitas diversas experiências com diferentes formas de rede neurais recorrentes.<sup>16</sup> As experiências iniciais foram feitas com o intuito de testar as ferramentas e tecnologias utilizadas, como também ter comprovação daquilo que a literatura estudada afirma, isto é, o desempenho superior das redes neurais recorrentes em detrimento das neurais normais e métodos estatísticos mais simples.

Os testes reais, já com suporte técnico e ferramental entendido e organizado, foram divididos em três fases:

- **Fase 1:** Treino com dados iniciais (3.1) sem uso de pré processamentos como média móvel e interpolação sobre janelas de dados faltantes. Contudo, com a inclusão de valores de fora do domínio (-1) sobre todo dado ausente.
- **Fase 2:** Treino com os mesmos dados iniciais da fase anterior, contudo com pré processamentos como média móvel e interpolação sobre janelas de dados faltantes. Adição de coluna extra de sazonalidade média (3.4).
- **Fase 3:** Fase final, com todos os procedimentos mencionados na *fase 2*, entretanto, contando com a adição de novos dados: meteorológicos (3.3) e socioeconômicos adicionais. (3.4).

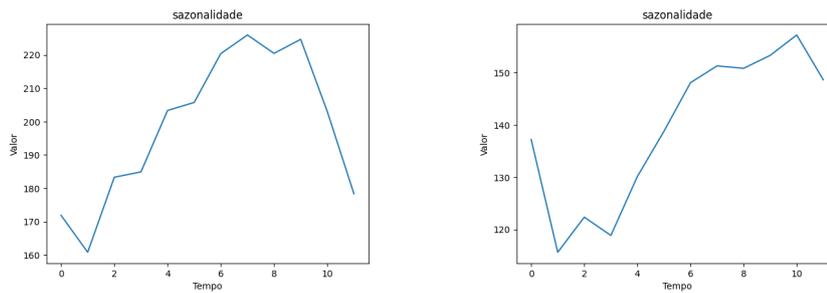
### 3.6.1 Pré treinamento

Após o pré-processamento dos dados, antes do início do treinamento, foram realizados procedimentos para analisar a correlação entre as variáveis por meio de uma matriz de correlação. Durante essa análise, identificamos e descartamos tipos de entrada que apresentavam forte correlação com um grande número de outras variáveis<sup>17</sup>.

---

<sup>16</sup> Esse processo empírico é natural na busca modelos de melhor desempenho (*CHUNG et al., 2014*) e é por si só uma área de estudo.

<sup>17</sup> Foi utilizado, para isso, o Fator de Inflação de Variância (VIF - *Variance Inflation Factor*). Optamos por um critério de limite de VIF elevado para o descarte, devido as redes neurais complexas tenderem a não serem muito impactadas por esses problemas.



(a) Sazonalidade média do estado de Goiás. (b) Sazonalidade média do estado do Pará.

**Figura 3.3:** Exemplo média de sazonalidade dos dados de treino.

Antes dos dados de entrada serem separados em dados de treino, dados de validação de treino e dados de teste de validação final, aplicamos um pré processamento de transformação de normalização ou padronização (2.2.4), sendo a escolha de cada qual sendo tratada como um hiperparâmetro a ser julgado pelos resultados dos erros quadráticos sobre os corpos de validação de treino.

Em sequência, foram separados os dados de entrada em dados de treino, que compreendem o intervalo do início do ano 2003 até o fim do ano de 2016. Contudo, uma fração<sup>18</sup> desse conjunto foi separada para validação de treino, e não é utilizada para configuração dos *pesos* e *bias* do modelo preditivo, mas sim como referência para os ajustes de hiperparâmetros.

Por fim, os dados referentes ao ano de 2017 foram deixados como teste final do modelo para previsão do consumo de cimento durante os meses de 2019 devido ao recuo dos dados de entrada (3.1.2). Acrescenta-se ainda que, essa escolha por um corpo de prova de teste inédito que não participou tanto da fase de treinamento quanto da de configuração de hiperparâmetros, é vital para reproduzirmos uma situação realista de predição onde não sabemos nenhuma informação do futuro. Deste modo, os resultados apresentados na seção de resultados (4) seriam próprios para uso pela indústria durante o ano de 2018.

### 3.6.2 Treinamento

Como referido, na seção anterior, cada experimento contou com uma amostra de validação de treino<sup>19</sup> que em cada iteração (*epoch*) tem seu erro quadrado calculado e monitorado, dessa forma podemos ajustar os hiperparâmetros e evitarmos sobreajuste (2.5.1).

Outra prática utilizada que desfavorece a ocorrência de sobreajuste é a de início de monitoramento atrasado, que pode ser delimitado com um valor arbitrário de modo a permitir o modelo maturar em seu desenvolvimento<sup>20</sup> ignorando o número de iterações correspondente ao valor delimitado. Quando monitoramos o modelo sem essa configuração

<sup>18</sup> Geralmente, uma parcela equivalente a um ano ou dois, sendo a delimitação dessa escolha também uma forma de configuração de hiperparâmetro.

<sup>19</sup> No TensorFlow, o erro correspondente a essa amostra de validação para treino é denominado “*Validation Loss*” (Erro de validação) [https://www.tensorflow.org/tutorials/keras/overfit\\_and\\_underfit](https://www.tensorflow.org/tutorials/keras/overfit_and_underfit)

<sup>20</sup> Na biblioteca TensorFlow, o referido tempo de maturação de monitoramento é delimitado pelo parâmetro

e resgatamos seu melhor estado, isto é, a iteração que obteve o menor valor de erro para o corpo de validação de treino, há o risco dessa iteração ser precoce, ocasionando um modelo de pouca complexidade e, muito provavelmente, com sobreajuste.

Devido a natureza estocástica do gradiente descendente escolhido (2.6), foi selecionado o modelo com melhor performance sobre os dados de validação (amostra interna) de treino entre um conjunto de dez modelos treinados. Além disso, sempre houve uma análise empírica sobre o resultado geral desses dez modelos para ajuste de hiperparâmetros.

### 3.6.3 Validação

Após, o treinamento ser concluído, seja por término de um limite de epochs arbitrário ou devido a não haver melhora na diminuição do erro monitorado em um intervalo de tempo também arbitrário<sup>21</sup>, coletamos o erro absoluto da diferença da previsão do modelo para o consumo de cimento referente do mês em questão. Esses erros, por sua vez, foram agrupados de modo a gerar as estatísticas de MAE, MAPE e RSME sobre o conjunto de valores previstos por estado.

Importa frisar que, apesar da opção de não ajustar os hiperparâmetros sobre resultados gerados pelos dados de entrada de teste final, isto é, os referentes ao ano de 2017 para se prever 2019, como relatado em (3.6.1), sempre repetimos treinamentos que resultavam em valores muito próximos de zero ou negativos para algum mês. Isso devido ao fato de, mesmo simulando o desconhecimento do futuro para o treinamento dos modelos, sabemos que resultados próximos de zero são improváveis e, portanto, *outliers* e os negativos são absurdos por não pertencerem ao domínio em questão.

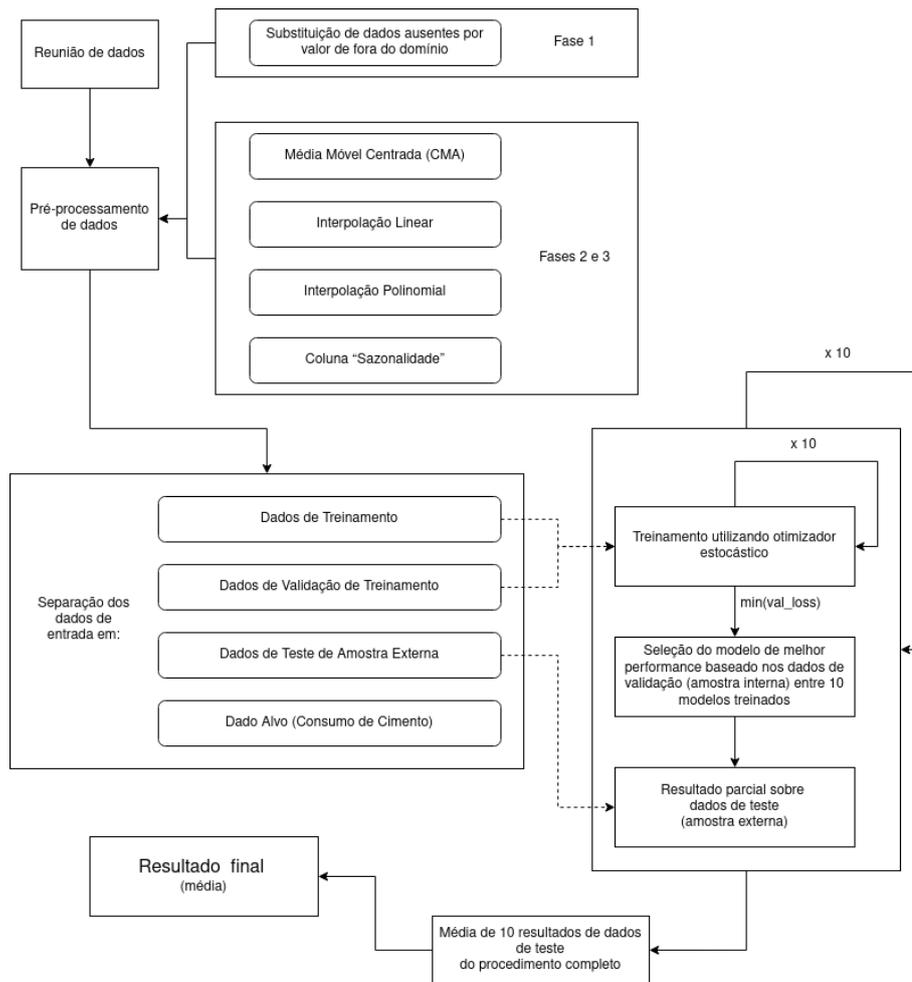
Após a escolha do modelo de melhor performance sobre os dados de validação, o mesmo é submetido a entrada relativa ao teste final (previsão de 2019 usando os dados de entrada de 2017 que são inéditos para o modelo) esse resultado final é coletado e todo o processo é repetido novamente dez vezes para que tenhamos uma validade estatística e *outliers* sejam diluídos em uma amostra média mais ampla.

---

*start\_from\_epoch* do método provedor de monitoramento *EarlyStopping*

<sup>21</sup> Interromper o treinamento devido não haver melhorar no erro é um pratica denominada “*Early stopping*” (Parada antecipada) e também favorece a não ocorrência de sobreajuste [https://www.tensorflow.org/api\\_docs/python/tf/keras/callbacks/EarlyStopping](https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/EarlyStopping)

## 3.7 | TECNOLOGIAS UTILIZADAS



**Figura 3.4:** Fluxo do treinamento.

O resultado final médio, por sua vez, tem seus erros calculados para a geração das estatísticas mencionadas, e toda sua informação salva em uma tabela de registro para processamento de apresentação (gráficos e tabelas).

### 3.7 Tecnologias utilizadas

Neste trabalho<sup>22</sup> foi utilizada uma coleção de software bem comuns no campo da ciência de dados e modelagem baseada em aprendizado de máquina. Abaixo uma lista com as principais:

<sup>22</sup> Os arquivos de código que constituíram este trabalho estão disponíveis em <https://github.com/Pedro84skynet/TCC-MAC0499>.

<b>Tecnologia</b>	<b>Descrição</b>
<b>Python:</b>	Uma das linguagens de alto nível mais utilizadas devido sua elevada versatilidade e acessibilidade. Sua gama de bibliotecas auxiliares a coloca entre as principais linguagem para trabalhos com ciência de dados. ( <i>Python Programming Language 2023</i> )
<b>NumPy:</b>	Biblioteca veloz, por ser escrita em C, e fundamental para computações numéricas complexas em Python. Oferece suporte a arrays multidimensionais e diversas funções matemáticas. ( <i>NumPy: Numerical Python 2022</i> )
<b>Pandas:</b>	Principal biblioteca para manipulação e análise de dados, facilitando operações com seus dataframes que são estruturas de dados tabulares. ( <i>Pandas: Python Data Analysis Library 2023</i> )
<b>Jupyter Notebook:</b>	Ambiente interativo para desenvolvimento de código, muito adequada para quando necessitamos de feedback visual ágil e experimentações. ( <i>Jupyter Notebook 2023</i> )
<b>TensorFlow:</b>	Um dos principais softwares de código aberto usado para o desenvolvimento de modelos de aprendizado de máquina e redes neurais. Muito versátil, com uma ampla coleção de métodos para diversos tipos de modelos e casos. ( <i>TensorFlow: An Open Source Machine Learning Framework for Everyone 2023</i> )
<b>SHAP:</b>	Ferramenta relativamente nova desenvolvida para o estudo de explicabilidade de modelos de aprendizado de máquina. Seu principal uso se formaliza nas estimativas de valor Shapley, calculando a contribuição de cada tipo de entrada para a predição. ( <i>SHAP (SHapley Additive exPlanations) 2023</i> )
<b>Matplotlib:</b>	Biblioteca gráfica de visualização de dados em Python, oferecendo diversas construções visuais como gráficos e plots. ( <i>Matplotlib: Visualization with Python 2023</i> )
<b>Statsmodels:</b>	Biblioteca estatística em Python muito usada para estimar erros de modelos, incluindo regressões, e análises de séries temporais. ( <i>Statsmodels: Statistical Models in Python 2023</i> )
<b>LibreOffice Calc:</b>	Planilha eletrônica de código aberto com recursos para cálculos e análise de dados. ( <i>LibreOffice Calc 2023</i> )

**Tabela 3.7:** Descrição das Tecnologias Utilizadas

# Capítulo 4

## Experimentos e Resultados

Todos os experimentos foram feitos com o intuito de melhorar o desempenho dos modelos gradualmente sem desprezar as boas práticas apresentadas em (2.6) e independentemente do tipo de entrada, foram feitos por unidade da federação com seus próprios ajustes de hiperparâmetros e análises de correlação.

O trabalho de ajustes de hiperparâmetros foi feito de forma empírica sempre avaliando o desempenho do modelo sobre os resultados de erro quadrados sobre o corpo de validação. Importante ressaltar que os resultados apresentados nessa seção não fazem parte desse corpo de validação. São resultados usando entradas inéditas, referentes ao ano de 2017<sup>1</sup>, não inseridas durante o treinamento e ajustes de hiperparâmetros.

### 4.1 Treinamento de modelos de predição

Para cada unidade da federação foram feitos diferentes experimentos com todos os tipos de dados apresentados em (3.2.1) divididos em três grupos de experimentos (3.6): os referentes aos dados denominados iniciais, brutos, sem aplicação da maioria dos métodos de pré-processamento apresentados; os referentes aos dados iniciais, mas, desta vez, pré processados e, finalmente, os treinos completos, com todos os dados (iniciais, meteorológicos e socioeconômicos adicionais) também pré-processados.

Foi decidido focar a análise das séries temporais geradas, para aprofundamento, nos cinco estados de maior consumo de cimento de cada região brasileira: Pará (Norte); Bahia (Nordeste); Goiás (Centro-oeste); São Paulo (Sudeste) e Paraná (Sul). Analisando, deste modo, as divergências de valor registrado e previsto para as séries temporais de consumo de cimento do ano de 2019.

---

<sup>1</sup> Como um detalhe técnico, as redes LSTM, na maioria dos casos, recebem seus dados de entrada como uma sequência, *timesteps*. Neste trabalho, foi usado como entrada uma janela de 36 meses, então parte da entrada, na realidade, não é inédita ao modelo, contudo a parte final dessa sequência, referente aos 12 meses recuados em dois anos, é sim inédita.

### 4.1.1 Fase 1: Treinamento sobre dados iniciais sem pré processamento

Essa bateria de treinamentos contou com todos os dados da tabela (3.1) em seu estado bruto, o único método de pré processamento usado foi o referente a imputação sobre os dados ausentes por valores de fora do domínio (2.2.1) de valor  $-1$ .

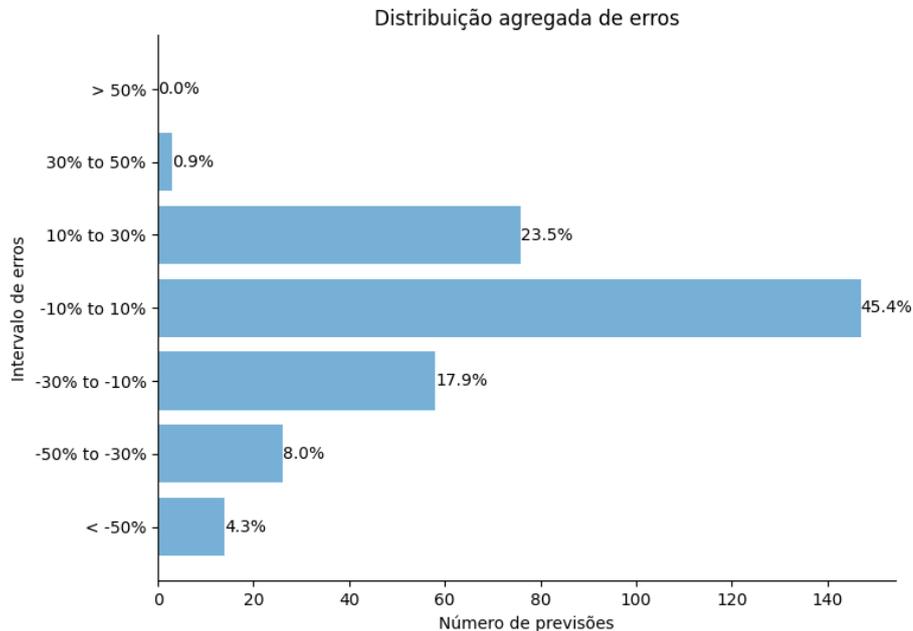
Estado	RMSE (mil ton.)	MAE (mil ton.)	MAPE %	Desv.Padrão <sup>a</sup> %
Acre	2,67	2,20	15,84	8,36
Alagoas	4,72	3,99	11,89	7,31
Amapá	2,75	2,27	23,60	16,07
Amazonas	7,23	5,93	12,43	7,86
Bahia	29,03	23,54	8,67	6,76
Ceara	32,61	30,58	22,24	9,76
Distrito Federal	8,15	7,01	12,51	6,46
Espirito Santo	14,83	12,17	13,85	12,00
Goiais	22,03	17,79	8,63	6,26
Maranhão	23,14	22,02	36,85	18,04
Mato Grosso	10,82	7,86	7,12	8,13
Mato Grosso do Sul	8,13	5,70	8,49	10,61
Minas Gerais	69,89	59,39	10,99	6,55
Pará	59,66	55,74	51,96	20,38
Paraíba	11,00	8,49	12,57	9,38
Paraná	31,12	21,81	7,65	9,87
Pernambuco	18,50	15,70	10,98	6,22
Piauí	10,43	8,68	16,87	15,06
Rio Grande do Sul	10,02	8,17	3,56	2,71
Rio de Janeiro	19,12	16,09	7,32	5,51
Rio Grande do Norte	8,57	7,42	14,88	7,01
Rondônia	4,32	3,42	11,24	10,58
Roraima	2,31	2,10	20,34	9,02
Santa Catarina	39,43	33,56	11,84	6,77
São Paulo	97,26	77,94	9,65	8,86
Sergipe	5,88	4,45	11,93	9,31
Tocantins	13,58	12,40	37,49	19,17
<b>Total</b>	<b>21,01</b>	<b>17,65</b>	<b>15,61</b>	<b>9,78</b>

**Tabela 4.1:** Resultados estatísticos do experimento com dados brutos.

<sup>a</sup> Desvio padrão sobre erros percentuais absolutos.

Os resultados acima mostram um esforço de predição com erro absoluto médio percentual de **15,61%** para as predições mensais de todos os estados no ano de 2019. Ao analisar a coluna do MAPE, fica evidente que os estados da região Norte e Nordeste apresentaram os piores resultados, destacando-se Amapá, Ceará, Maranhão, Pará e Tocantins como os cinco com o maior erro. No entanto, é crucial notar que esses estados têm populações menores, o que minimiza o impacto desse erro em termos de volume de cimento. Por exemplo, mesmo

com um erro percentual médio de 23.60%, o Amapá registrou apenas 2.27 mil toneladas de erro médio absoluto. Em contraste, Minas Gerais, com um MAPE de 10,99%, apresentou um erro absoluto médio de 59,39 mil toneladas.



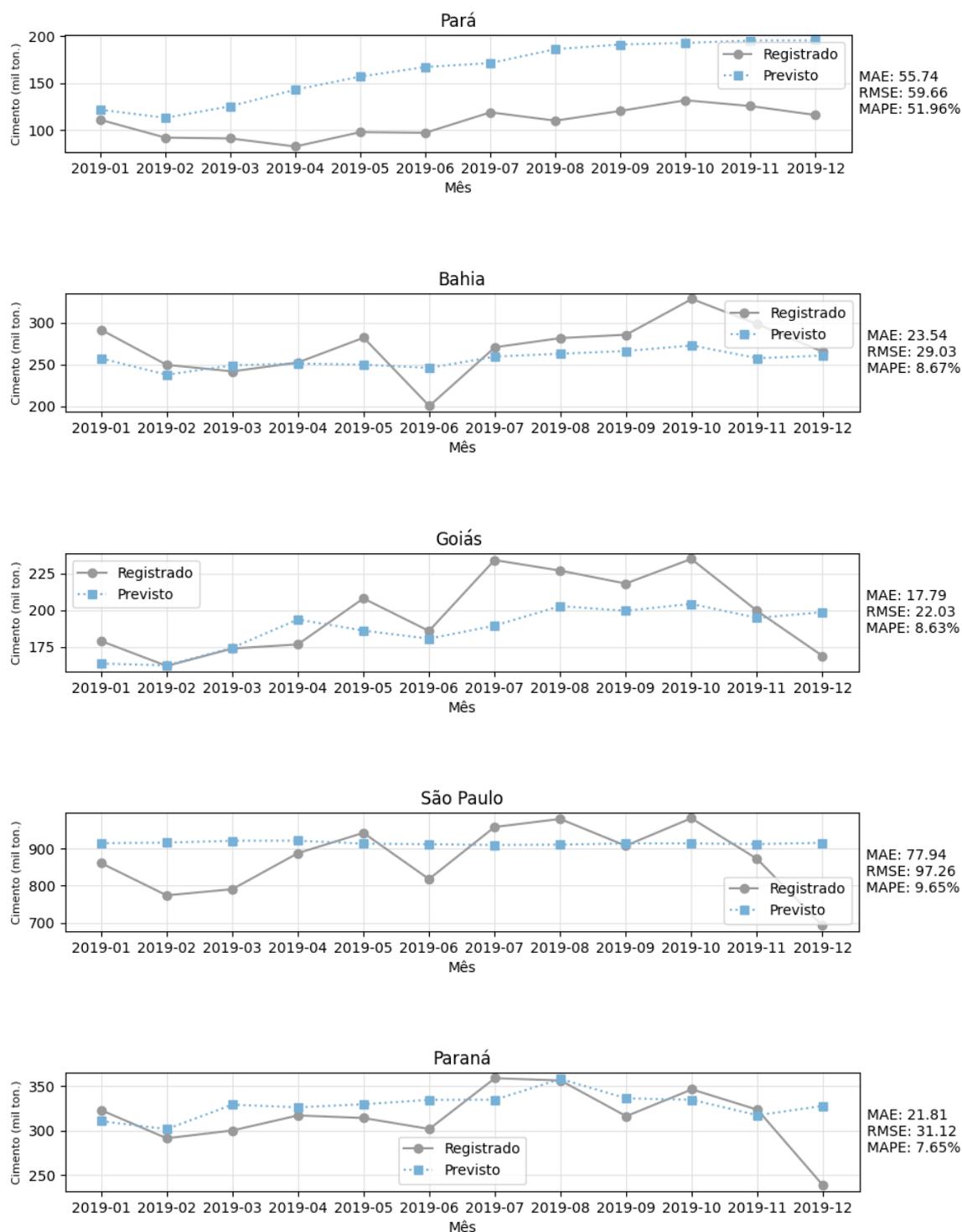
**Figura 4.1:** Distribuição agregada dos erros dos dados iniciais brutos.

A análise da distribuição dos erros para os dados iniciais brutos revela cerca de 45.4% dos erros concentrados numa faixa estreita de -10% a 10%, indicando uma precisão considerável nas previsões que se alinham muito próximas aos valores reais. Isso mostra que a escolha desses dados estatísticos tradicionais (3.1) continua válida e que a organização da estrutura da rede neural teve efeito.

Ademais, observamos que 23.5% dos erros estão entre 10% e 30%, enquanto 17,9% estão na faixa de -30% a -10%, destacando uma distribuição de estimativas superestimadas maior, entretanto, a soma das faixa de erros menores que -10% mostram em torno do zero. No entanto, é relevante observar que a soma dos erros das faixas -30% a -10%, -50% a -30% e menores que -50%, deixam claro um cenário de leve subestimação do consumo real.

Nos gráficos de séries temporais previstas (4.2), vemos que o estado do Pará, o MAPE de 51,96% indica uma divergência discrepante entre as previsões e os valores reais com uma elevada superestimativa. Os estados da Bahia, com um MAPE de 8,67%, e São Paulo, com MAPE 9,65% apresentam previsões mais assertivas, contudo não capturaram a sazonalidade daquele ano.

Em Goiás, o MAPE de 8,63% mostra uma aproximação razoável, percebemos que o modelo conseguiu capturar um pouco melhor às tendências sazonais ao longo do ano, embora tenha ocorrido uma subestimação do valor real no segundo semestre. O estado do Paraná com um MAPE de 7,65%, que mostrou-se o melhor resultado dentre os estados selecionados, teve um bom acompanhamento da sinuosidades do dado real, exceto para o mês de dezembro.



**Figura 4.2:** Resultados das previsões de cimento para 2019 - Dados iniciais.

### 4.1.2 Fase 2: Treinamentos sobre dados iniciais com imputação de interpolação e média móvel

Este treinamento, contou com praticamente tudo que fora apresentado neste trabalho sobre pré processamento dos dados no sentido de imputação e transformação de dados. Os experimentos consistiram em aplicações dos pré processamentos de média móvel e interpolação (linear ou polinomial) para as janelas de dados faltantes<sup>2</sup>; criação de uma coluna de dados de sazonalidade média cíclica (3.4) denominada sazonalidade média cíclica (3.4) e análise e exclusão de poucos dados baseada em mapa de correlação.

Estado	RMSE (mil ton.)	MAE (mil ton.)	MAPE %	Desv.Padrão <sup>a</sup> %
Acre	2.12	1.82	15.78	13,57
Alagoas	4.93	3.66	11.43	12,20
Amapá	2.30	1.92	22.57	19,38
Amazonas	5.76	4.81	11.28	8,92
Bahia	42.93	36.85	13.41	7,87
Ceara	24.26	18.52	15.03	15,38
Distrito Federal	7.86	6.56	12.87	10,14
Espirito Santo	9.76	8.00	8.80	7,23
Goiais	23.92	18.48	8.73	6,81
Maranhão	15.78	13.34	24.10	19,80
Mato Grosso	24.55	19.07	14.25	10,64
Mato Grosso do Sul	11.96	8.63	12.79	15,14
Minas Gerais	67.49	54.80	10.07	6,58
Pará	23.46	18.86	19.75	17,25
Paraíba	7.81	6.66	10.91	8,27
Paraná	31.12	22.64	7.67	8,97
Pernambuco	19.79	14.44	9.70	8,16
Piauí	10.05	8.65	16.63	13,96
Rio Grande do Sul	9.80	6.93	3.05	3,26
Rio de Janeiro	37.59	31.31	13.16	8,43
Rio Grande do Norte	8.17	6.34	15.42	14,92
Rondônia	5.04	4.25	12.78	8,75
Roraima	2.56	2.43	23.45	7,57
Santa Catarina	46.60	36.83	12.81	9,77
São Paulo	95.16	83.04	9.55	5,84
Sergipe	6.24	5.06	14.32	9,91
Tocantins	4.83	3.73	12.00	11,20
<b>Total</b>	<b>20.44</b>	<b>16.58</b>	<b>13.42</b>	<b>10,74</b>

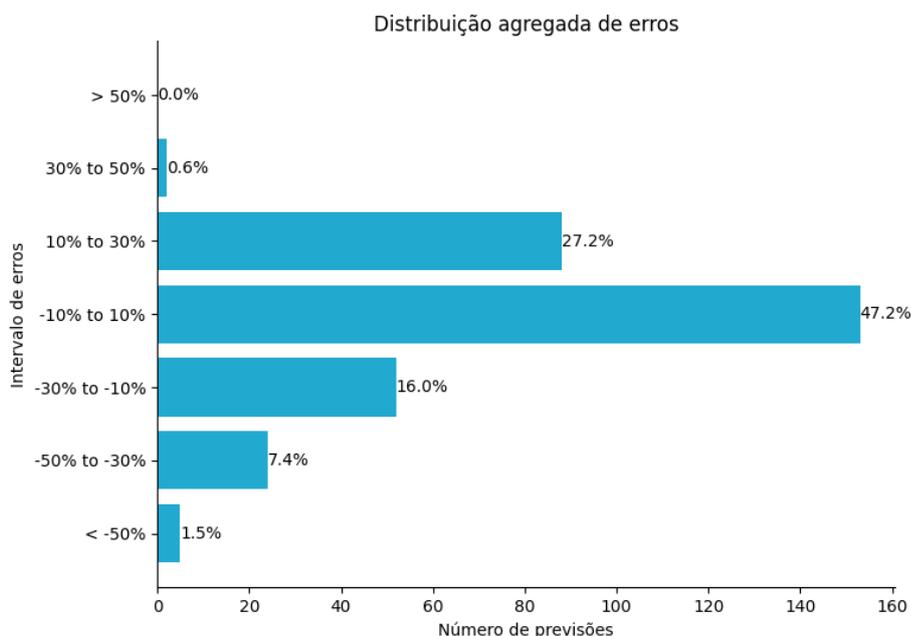
**Tabela 4.2:** Resultados estatísticos do experimento com dados iniciais com pré-processamento adicional.

<sup>a</sup> Desvio padrão sobre erros percentuais absolutos.

Essa bateria de testes demonstrou melhora no quadro geral da assertividade dos modelos de predição representado por uma queda do MAPE total de 15,61%, relativo a fase 1 (4.1.1)

<sup>2</sup> Relação dos métodos usados na tabela (3.2)

para 13,42%. Contudo, apesar de termos melhoras significativas, em destaque para os estados do Pará e Tocantins, que diminuíram seus erros absolutos percentuais médios de 51,96% para 19,75% e 37,49% para 12,00%, respectivamente, alguns estados tiveram piora como Bahia, Roraima, Mato Grosso do Sul e Paraná, Rondônia e, em especial, Mato Grosso que teve um crescimento de 7,13% de seu MAPE comparado ao resultado anterior.



**Figura 4.3:** Distribuição agregada dos erros dos dados iniciais com pré-processamento adicional.

A distribuição agregada dos erros medidos nessa fase mostra, principalmente, uma melhora na faixa dos erros de -10% a 10% passando a ter 47,2% dos erros incluídos na mesma, em contraste a 45,4% da fase anterior. Também tivemos uma diminuição das faixas relativas a grandes erros como os menores de 50% do valor registrado, passando a ter somente 1,5% dos erros nessa faixa em relação a 4,3% da fase anterior,

O resultado mantém, de certo modo, a mesma relação entre as proporções de erros agregados relativos à superestimação e subestimação em relação à fase anterior (4.1.1). Com um maior número de erros agregados na faixa de 10% a 30%, contudo, desta vez, a soma das faixas de erros relativos a subestimação é menor. Portanto, houve um cenário de leve superestimação das previsões com 27,2% dos erros na faixa de 10% a 30% em contraposição a somente 16,0% deles na faixa de 30% a -10%, uma queda de 0,5% dos erros na faixa de -50% a -30%.

Analisando a série temporal de predição gerada nessa fase (4.4) vemos uma aproximação média constatada pelos dados estatísticos. Entretanto, vemos alguns pontos importantes a serem ressaltados quando comparamos com as séries da fase anterior (4.2). O estado do Pará que teve a maior melhora na assertividade, perdeu um pouco a captura da tendência<sup>3</sup> de crescimento entre o período de 2019-05 a 2019-10 e decrescimento do primeiro mês do ano.

<sup>3</sup> Tendência como atributo de séries temporais (2.1.1).

4.1 | TREINAMENTO DE MODELOS DE PREDIÇÃO

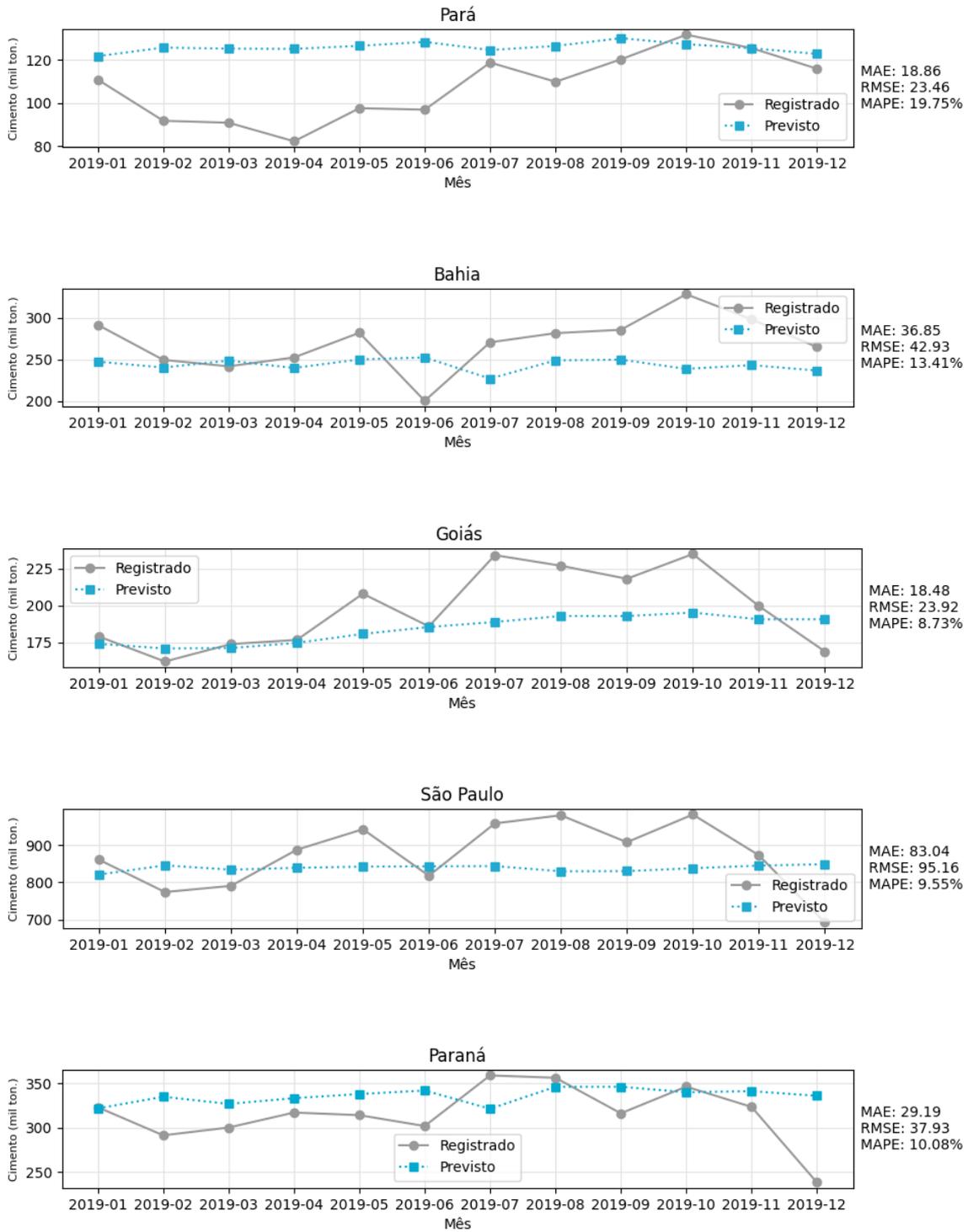


Figura 4.4: Resultados das previsões de cimento para 2019 - Dados iniciais Processados.

Os resultados de Bahia, São Paulo e Paraná, apesar de serem mais precisos que os anteriores ainda demonstram um fracasso do modelo em capturar as nuances sazonais do consumo de cimento. O estado de São Paulo, em particular, mostra uma aproximação de predição mais similar a uma regressão linear, ignorando completamente qualquer período inserido no ano em questão.

### 4.1.3 Fase 3: Treinamentos com dados socioeconômico adicionais e dados meteorológicos

O conjunto total de dados, inicial, meteorológicos e os adicionais, ou adicionais (3.4) consistiu em um vetor de entrada de aproximadamente<sup>4</sup> 36 tipos de entradas.

Essas entradas passaram pelos mesmos procedimentos de pré-processamento da fase anterior, ou seja, média móvel centrada e interpolação sobre dados faltantes. Assim como pouquíssimas mudanças em relação a ajustes de hiperparâmetros. Somente, houve necessidade dos ajustes de hiperparâmetros quando era constatado um aumento do erro muito elevado no conjunto de amostras de validação usada durante o treinamento<sup>5</sup> ou quando algum resultado era anormal segundo o que foi discutido em (3.6.3).

Essa fase final de treinamento de modelos preditivos com dados meteorológicos e socioeconômicos adicionais incluídos, apresentou o melhor resultado frente as fases anteriores como é observado pelos resultados da tabela (4.3). Foi obtido uma queda do RMSE, MAE, MAPE e seu desvio padrão, totais, para apenas 15,71%, 12,90%, **9,88%** e 7,32%, respectivamente. Essa queda geral, de todas métricas, é um importante indicativo de progresso, pois testemunha um modelo que melhorou a assertividade de suas predições sem apresentar inserção de ruídos, que poderiam ser acusados por uma não diminuição do RSME ou aumento de dispersão, representado por uma não diminuição do desvio padrão.

Em relação a fase anterior (4.1.2), tivemos melhorias significativas nas previsões da grande maioria dos estados das grandes regiões norte e nordeste, com destaque para Acre, Amapá, Maranhão, Pará, Piauí, Rondônia e Roraima que tiveram seu MAPE reduzido em 7,7%, 14,17%, 9,99%, 11,65%, 9,08%, 3,91% e 7,37%, respectivamente, uma melhoria de 63,87% no total para esses sete estados. Somente o estado do Tocantins teve uma piora considerável, passando de 12,0% de seu MAPE para 19,86%, e alguns outros estados como Bahia, Rio grande do Sul, Rio de Janeiro e São Paulo apresentaram ínfimas melhorias ou pioras em relação aos resultados da fase passada.

---

<sup>4</sup> O uso do termo "aproximadamente", é porque alguns tipos de dados de entrada não estão disponíveis para todos os estados. E.g.: produção de cimento que é ausente nos estados do Acre, Amapá e Roraima.

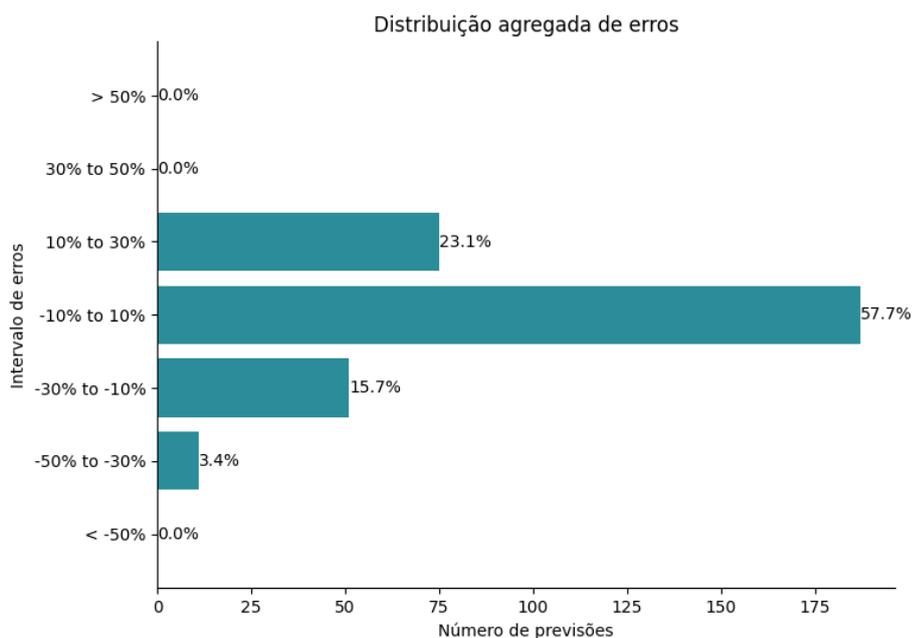
<sup>5</sup> Lembrando que, essa amostra de validação, não é a do teste relativo aos resultados referentes ao ano de 2019 apresentados nesta seção.

Estado	RMSE (mil ton.)	MAE (mil ton.)	MAPE %	Desv.Padrão <sup>a</sup> %
Acre	1.17	0.94	8.08	7,09
Alagoas	5.11	4.17	12.20	8,03
Amapá	1.09	0.73	8.40	9,64
Amazonas	6.38	5.19	10.90	7,06
Bahia	28.75	20.35	7.87	9,39
Ceara	14.66	13.10	9.38	5,08
Distrito Federal	7.78	5.63	11.06	12,08
Espirito Santo	14.46	10.54	12.10	13,45
Goiais	12.85	10.98	5.55	3,51
Maranhão	10.24	8.58	14.11	9,88
Mato Grosso	9.57	7.64	6.84	6,30
Mato Grosso do Sul	7.23	5.68	7.83	7,00
Minas Gerais	69.17	62.44	11.78	5,39
Pará	10.50	8.83	8.10	5,11
Paraíba	8.49	7.01	10.36	6,07
Paraná	36.17	28.94	9.80	9,23
Pernambuco	18.01	14.61	10.14	6,89
Piauí	4.75	3.84	7.55	7,38
Rio Grande do Sul	11.28	9.84	4.25	2,38
Rio de Janeiro	27.57	23.58	10.19	6,06
Rio Grande do Norte	5.22	4.64	9.70	4,41
Rondônia	3.74	2.75	8.87	9,62
Roraima	1.80	1.66	16.08	6,22
Santa Catarina	21.34	18.99	6.92	3,48
São Paulo	74.32	56.90	7.11	7,25
Sergipe	5.12	4.17	11.76	7,73
Tocantins	7.30	6.49	19.86	11,98
<b>Total</b>	<b>15.71</b>	<b>12.90</b>	<b>9.88</b>	<b>7,32</b>

**Tabela 4.3:** Resultados estatísticos do experimento com dados adicionais e com pré-processamento adicional.

<sup>a</sup> Desvio padrão sobre erros percentuais absolutos.

O gráfico de distribuição agregada dos erros (4.3) para essa fase final também revela a melhoria de performance geral dos modelos de aprendizado de máquina. Primeiramente, há o destaque para um aumento dos erros incluídos na faixa central de *-10% a 10%*, relativa a maior precisão, de 47,2%, da fase passada, para 57,7% nesta mesma. As faixas de erros imediatamente maiores e menores a última citada ilustram tendência a um quadro geral de sobrestima ao passo que seus correspondentes agregamentos, somam 23,1% de erros em *10% a 30%* em contraponto a soma de 19,1% dos erros correspondentes as faixas de *-30% a -10%* e *-50% a -30%*.



**Figura 4.5:** Distribuição agregada dos erros dos dados adicionais e com pré-processamento adicional.

Sobre os erros um pouco maiores, também, obtivemos diminuição sendo a faixa de 30% a 50% zerada e a correspondente à -50% a -30% reduzida de 7,4% para 3,4%. Nessa fase, conseguimos eliminar a presença de erros nas agregações que representam os maiores desvios, ou seja, a dos erros maiores que 50% e menores que -50%.

A análise da seleção das séries temporais, desta fase (4.6), apresentam interessante mudanças quando comparado as séries das fases anteriores.

4.1 | TREINAMENTO DE MODELOS DE PREDIÇÃO

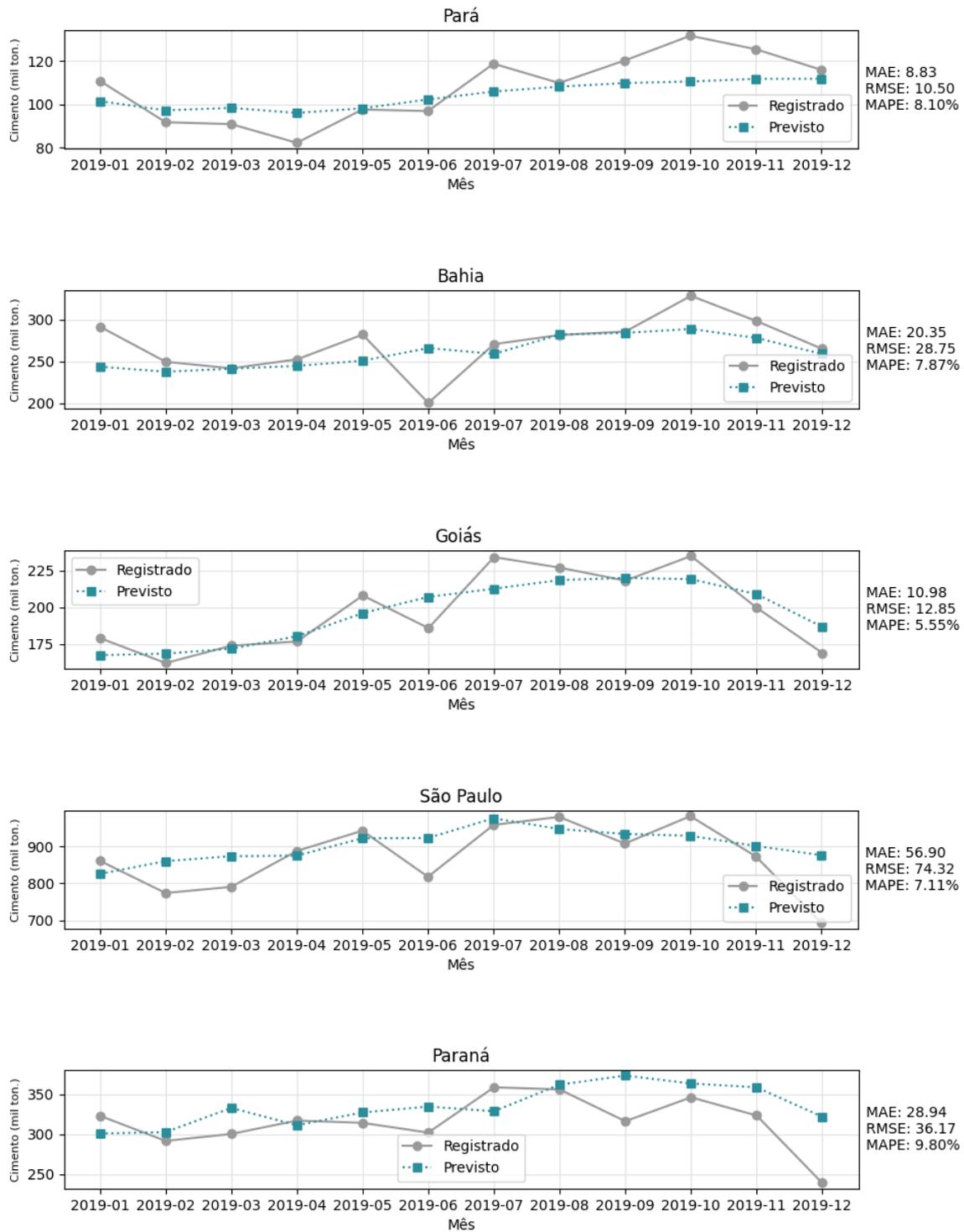


Figura 4.6: Resultados das previsões de cimento para 2019 - Dados adicionais.

No Estado do Pará, percebe-se uma leve captura da tendência de descida do primeiro semestre e crescimento no segundo, apesar de ter erros pontuais expressivos, teve uma grande melhora relativa a fase anterior (4.1.2) com seu MAPE tendo o melhor resultado dentre todas fases, isto é, 8,10%.

Os estados da Bahia e Paraná, demonstraram uma pequena captação da sazonalidade e tendências de crescimento e diminuição semestrais, embora ainda com erros significantes com destaque para o mês de junho para Bahia e dezembro para Paraná.

Goiás se destacou, não somente pelo seu MAPE de 5,58%, mas por melhor capturar a sazonalidade e a tendência do consumo daquele ano, com exceção de alguns ruídos em períodos específicos. São Paulo teve, nesta fase, seu melhor resultado indicado pelo MAPE de 7,25%, embora, seus modelos não tenham conseguido capturar a tendência de queda do início do ano e uma expressiva queda em junho.

## 4.2 Explicabilidade usando métodos SHAP

Como mencionado, em 4.1.1, para uma análise mais minuciosa, optamos por focar nos cinco estados de maior consumo de cimento de cada grande região brasileira. Deste modo realizamos o estudo de explicabilidade utilizando a biblioteca *SHAP (SHapley Additive exPlanations)* (2023). Essa biblioteca, calcula os valores SHAP referente a cada tipo de dado do vetor de entrada, como explicado sucintamente em (2.8.2). Essa análise é muito mais efetiva quando explicamos resultados que tiveram grande assertividade.

Cada gráfico do tipo "waterfall"<sup>6</sup>, da biblioteca referida, apresenta, elencados de cima para baixo em ordem decrescente, os prováveis valores de contribuição de cada tipo de entrada. São "prováveis" pois o cálculo determinístico dos valores SHAP é muito custoso computacionalmente, sendo necessários  $2^{n-1}$  cálculos de adição shap (2.28), com  $n$  sendo a dimensão do vetor de entrada, para cada fator. Sendo assim, são usadas operações estocásticas em um modelo chamado de *Kernel SHAP* (LUNDBERG e LEE, 2017). Por questões de praticidade, os dados de todas as entradas que menos contribuíram para o resultado em valor absoluto, fora as nove melhor contribuíram, são agregadas no final do gráfico como as  $n - 9$  outros tipos de entradas ( $n - 9$  other features).

$E[f(x)]$  é o valor padrão do modelo para uma entrada denominada de "plano de fundo". Devido ao fato dos modelos, aqui apresentados, terem sido submetidos por pré-processamento de padronização<sup>7</sup> (2.2.4) esse plano de fundo se formaliza como resultado do modelo quando todas as entradas são médias, ou seja, iguais a zero.  $f(x)$ , por sua vez, é o valor previsto pelo modelo para a determinada explicação apresentada.

Abaixo, elencamos alguns meses dos estados selecionados onde o modelo obteve boas previsões.

<sup>6</sup> Esses gráficos são gerados automaticamente pela biblioteca SHAP e fazem parte de uma coleção de tipos de gráficos pensados para auxiliar a leitura e interpretabilidade do usuário.

<sup>7</sup> Esses valores padronizados são os valores numéricos em cinza claro na frente dos nomes de cada tipo de entrada.

4.2 | EXPLICABILIDADE USANDO MÉTODOS SHAP

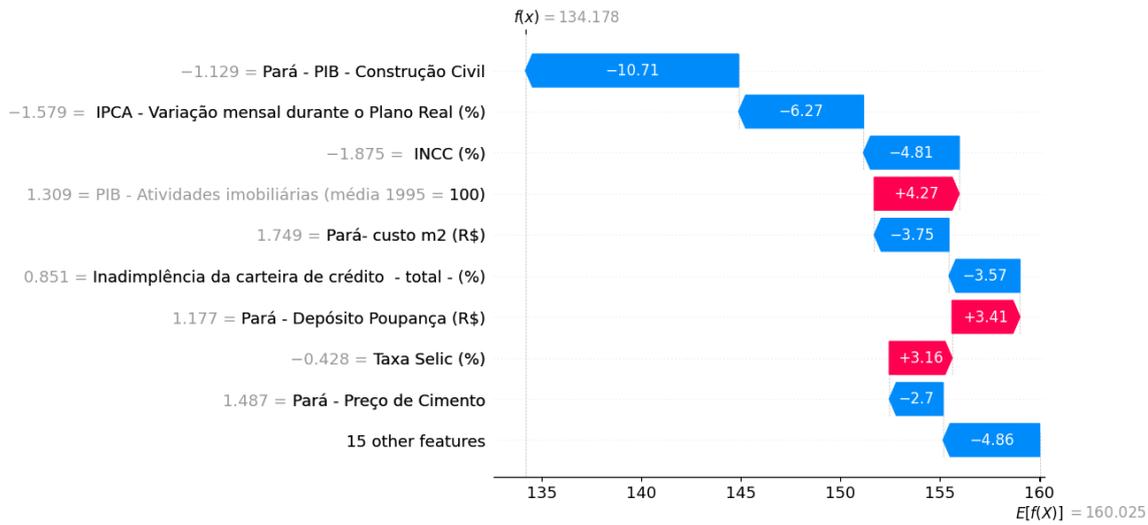


Figura 4.7: Explicabilidade waterfall para o estado do Pará.

O gráfico da explicabilidade da previsão do consumo de cimento do estado para o Pará, referente ao mês de outubro de 2019, teve erro absoluto de somente 1,94%, com uma previsão de consumo 134,17 mil toneladas versus um consumo registrado de 131,62 mil toneladas de cimento. Os fatores que melhor contribuíram esse resultado foram o PIB referente a construção civil, Índice Nacional de Preços ao Consumidor (IPCA) e Índice Nacional de Custo da Construção (INCC), todos agindo para a diminuição do valor médio  $E[f(x)]$ . Isso mostra um cenário ligado tanto ao poder de compra da população quanto a investimentos na construção civil, em geral.

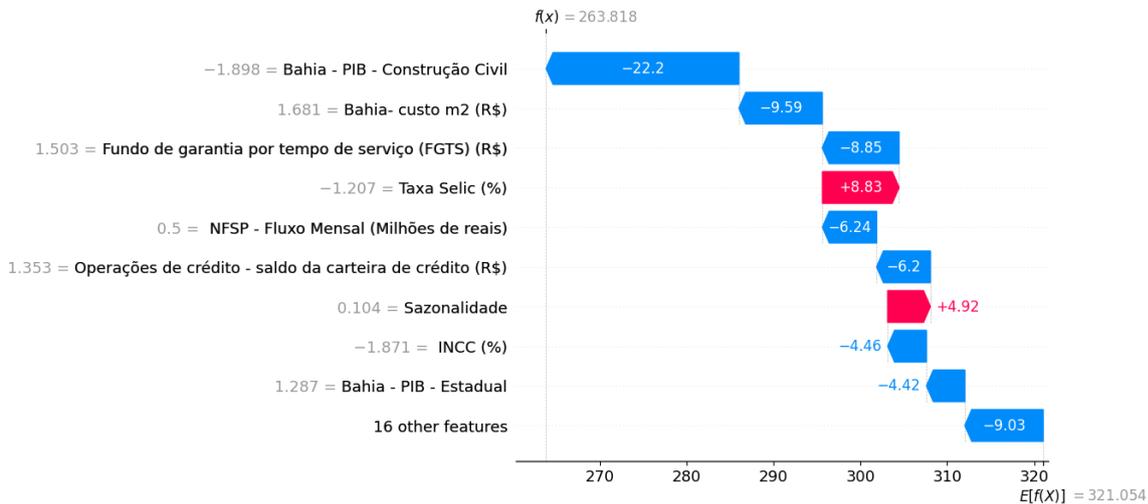
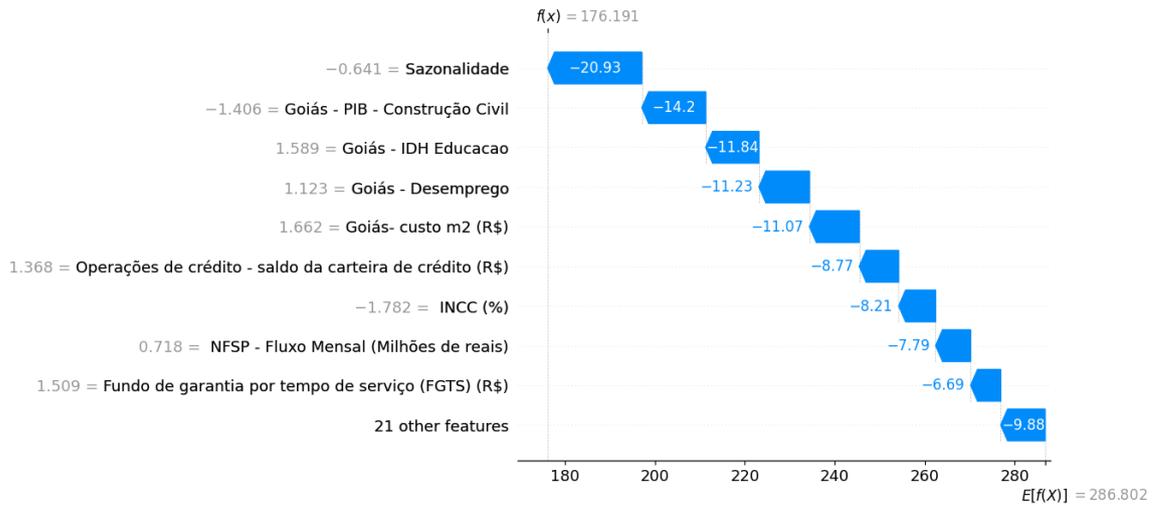


Figura 4.8: Explicabilidade waterfall para o estado do Bahia.

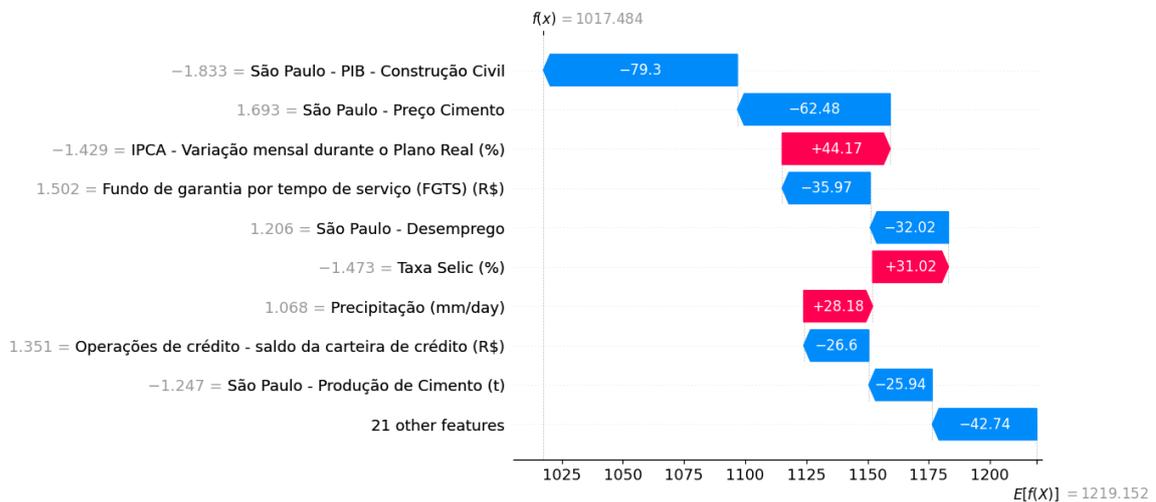
Para o estado da Bahia, destacamos o mês de julho, com uma previsão de 2,47% de erro absoluto percentual, sendo 263,81 mil toneladas previstas, em contraponto a 270,51 mil toneladas registradas. Também vemos aqui, como um índice de destaque, o PIB referente a construção civil, seguido pelo custo do  $m^2$  de área de construção e o FGTS. Taxa Selic foi o quarto possível fator que melhor contribuiu para previsão e ordem de valor absoluto, que

contribuiu para o aumento do valor resultante em oposição aos outros.



**Figura 4.9:** Explicabilidade waterfall para o estado do Goiás.

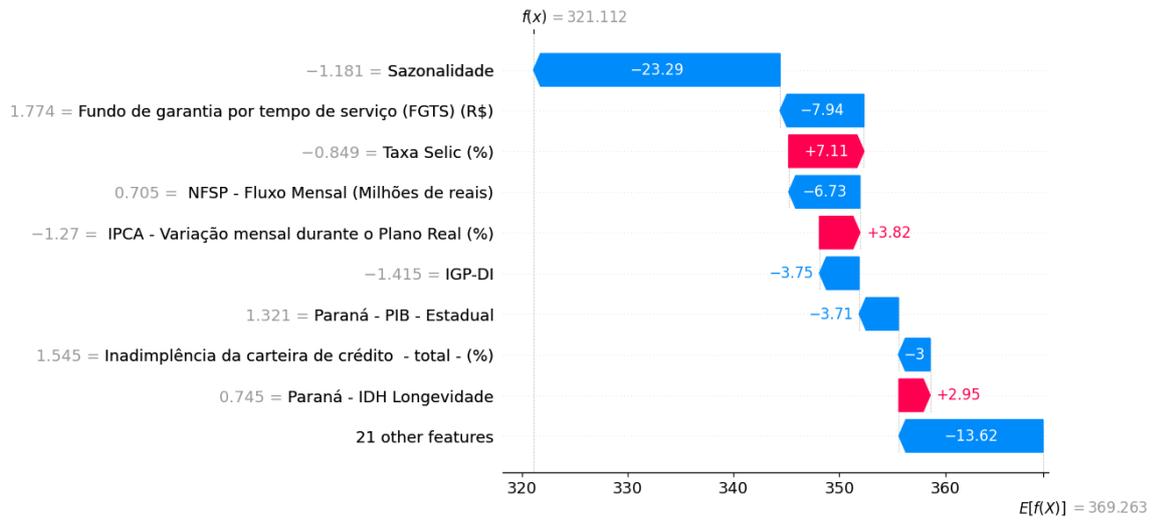
O erro absoluto percentual, para o estado de Goiás, no mês, de abril, foi de apenas 0,25%. Tivemos um resultado de previsão 176,19 vs. 176.62 mil toneladas de valor registrado. Aqui temos, uma interessante confirmação de fator de melhoria de performance, a coluna cíclica de sazonalidade média (3.4) introduzida nas fase 2, 3, muito provavelmente contribuí para esse bom resultado como é indicado no gráfico acima. Esse ponto também é destacado pela melhoria de assertividade visualizada no gráfico dos resultados da série temporal (4.6), relativo ao estado de Goiás, na fase 3.



**Figura 4.10:** Explicabilidade waterfall para o estado do São Paulo.

São Paulo teve sua melhor previsão no mês de agosto, com um erro absoluto percentual de 3,93% resultado da comparação de uma previsão de 1017,48 mil toneladas frente a um valor registrado de 978,99 mil toneladas para o referido mês. Novamente, temos como fator contribuinte protagonista o PIB da construção civil, seguido pelo preço do cimento e, novamente, IPCA. Indicando uma dinâmica também voltada para investimentos em construção civil e poder de compra dos consumidores.

## 4.2 | EXPLICABILIDADE USANDO MÉTODOS SHAP



**Figura 4.11:** *Explicabilidade waterfall para o estado do Paraná.*

Por fim, temos o gráfico *waterfall* para a explicabilidade do resultado de previsão de cimento para o estado do Paraná no mês de Janeiro de 2019. Tivemos um resultado de 321,11 mil toneladas para o valor previsto sendo, o registrado, 323,00 mil toneladas, ou seja, uma diferença de 0,58% de erro absoluto percentual. Aqui vemos, mais uma vez, um grande auxílio da coluna personalizada de sazonalidade média cíclica. Também temos, em destaque, os valores de FGTS e taxa Selic que novamente mostram-se como valores notáveis a se considerar em uma previsão de consumo de cimento.



## Capítulo 5

# Interpretação dos resultados

Conjunto de dados	RMSE (mil ton.)	MAE (mil ton.)	MAPE %	Desv.Padrão <sup>a</sup> %
Dados iniciais brutos	21,01	17,65	15,61	9,78
Dados com interpolação e média móvel	20,44	16,58	13,42	10,74
Novos dados inseridos	<b>15,71</b>	<b>12,90</b>	<b>9,88</b>	<b>7,32</b>

**Tabela 5.1:** Tabela de resultados totais.

<sup>a</sup> Desvio padrão sobre erros percentuais absolutos.

Esses resultados apresentados na tabela 5.1 mostram tendência de aprimoramento em relação às métricas estatísticas utilizadas para avaliar a precisão dos modelos de previsão de demanda de cimento ao passo em que avançamos nas etapas de pré-processamentos adicionais e inclusão de novos dados.

Utilizando os dados na sua forma bruta, sem pré-processamento de imputação avançado e média móvel observamos um desempenho razoável dos modelos, com um RMSE de 21,01 mil ton., MAE de 18,62 mil ton. e MAPE de 15,75%. Contudo, ao aplicar essas técnicas sobre os dados faltantes, é notado uma melhoria em todas as métricas. Os dados pré processados inicialmente reduziram o RMSE para 20,44 mil ton., MAE para 16,58 mil ton. e o MAPE para 13,42%, indicando uma maior precisão na previsão. Apesar que leve aumento de quase 1% do desvio padrão do MAPE geral.

Além disso, ao expandir o escopo dos dados de entrada adicionando mais dados, houve melhorias ainda mais expressivas. Esse último teste com os dados da tabela (3.3) e (3.4) adicionados, revela uma redução adicional do RMSE para 15,71, do MAE para 12,90 e do MAPE para 9,88. Esses resultados destacam uma tendência de aprimoramento na forma de diminuição dos erros de previsão, refletindo uma maior acurácia e robustez do conjunto de resultados dos modelos de previsão de séries temporais com a incorporação dos procedimentos citados.

Em geral, não foi alcançado um resultado satisfatório quanto a captura dos atributos de variação sazonal (2.1.1) da série temporal prevista. Isso ocorreu, mesmo com a contínua diminuições dos indicadores estatísticos de erros usados. Contudo na grande maioria dos

casos referentes a fase 3 tiveram boas representações das tendências semestrais durante o ano.

Não menos importante, houve uma grande e progressiva melhora no dificultoso território de previsão de consumo de cimento que se formaliza nas regiões do norte e nordeste do país. Com estados como o Amapá, Pará e Maranhão tendo seu erros de MAPE iniciais de 23,60%, 51,96% e 36,85% diminuídos para 8.40%, 8.10% e 14.11% respectivamente.

A investigação dos resultados das contribuições das variáveis de entrada, através da análise aditiva dos valores SHAP, desempenhou um papel significativo na elucidação do cenário subjacente ao problema da previsão do consumo de cimento. No ranqueamento gerado pelo gráficos *waterfall*, podemos ver recorrentemente presentes variáveis como as referentes a construção, PIB ligado a construção civil, custo do metro quadrado de área construída, assim como aquelas de esfera mais econômica, como FGTS, taxa SELIC e, em segundo plano, NFSP. Além disso, variáveis ligadas ao poder de compra da população, como operações de crédito, inadimplência e IPCA, também se destacam. Esse panorama ilustra um cenário que corresponde às expectativas de um consumo intimamente ligado às condições financeiras do país, aos investimentos no setor e ao poder de compra da população em geral.

Observa-se ainda que a coluna gerada pelas informações de sazonalidade de cada estado tem um impacto significativo em regiões onde a indústria cimenteira ainda não está totalmente consolidada, ou onde a população possui práticas culturais bem definidas que resultam em grandes variações na demanda ao longo do ano.

## 5.1 Possíveis próximos passos para futuros trabalhos

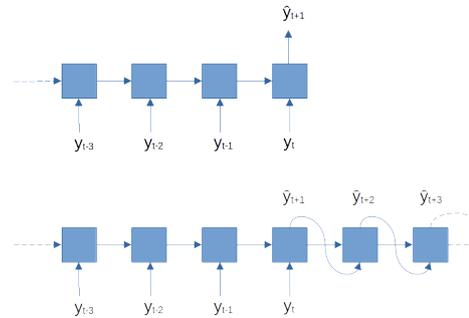
O trabalho de previsão de valores em séries temporais é um trabalho contínuo. Ao passo que novos dados e metodologias surgem, sempre há espaço para melhorias e adaptações. Com esse trabalho essa invariante permanece e são diversos os vetores de ação para, tanto ampliar as possibilidades preditivas, quanto alcançar ainda maior assertividade.

O principal, caminho, como é o caso da maioria dos estudos de caráter preditivo, é buscar novos tipos de dados fortemente representativos da relação entre consumidor e produto. Uma análise minuciosa desses dados para um preciso recuo, baseado em correlação, feito para cada variável, assim como novas modalidades de pré-processamento são sempre possibilidades ricas.

O ajuste fino dos hiperparâmetros com técnicas mais sofisticadas de AutoML, como a otimização Bayesiana e otimizações de multi-fidelidade, que procuram evitar o problema de alto custo da busca de grade para permutações de elevado número, podem ser chave para um ajuste ainda mais fino.

A metodologia empregada para produção dos resultados apresentados neste trabalho é totalmente aplicável, em presente ano, para previsão de consumo de cimento, por unidades federativas brasileiras, do ano seguinte, assim que houver disponibilidade de todos os dados públicos apresentados referentes ao ano passado. Contudo, informações de alta

assertividade histórica sobre o futuro dificilmente são dispensáveis. Tendo isso em mente, um próximo trabalho poderia ampliar a ambição de janela de predição para que no presente fosse possível prever, com uma precisão aceitável, uma janela maior que de um ano de dados do futuro usando recursos como modelagem autoregressiva.



**Figura 5.1:** Exemplo de arquitetura autoregressiva

Outro ponto é que, devido a limites inerentes de tempo e necessidade de aprofundamento de conhecimento de uso da biblioteca *SHAP (SHapley Additive exPlanations)* (2023), como mencionado em (3.1), foi decidido fazer sua análise sobre um conjunto treinado por redes neurais feed-forward. Contudo, seus resultados, apesar de serem comparáveis não são tão assertivos quanto os gerados pelas redes LSTM. Outra ação, seria o uso do instrumento de explicabilidade nessas redes de maior complexidade, além de sua análise não ser limitada a eventos pontuais.



# Capítulo 6

## Conclusão final

A investigação realizada sobre a previsão de séries temporais do consumo de cimento nos estados do Brasil revelou os aspectos intrínsecos e desafios complexos nessa área de atuação dos modelos de aprendizado de máquina. Essas previsões são fundamentais para a indústria, pois influenciam diretamente decisões estratégicas.

Durante esta pesquisa, foram explorados os mecanismos de aprendizado de máquina, enfatizando técnicas vitais de pré-processamento, como recuo de entrada, interpolação linear e polinomial, transformações e média móvel centrada. Priorizamos o uso de redes neurais recorrentes do tipo LSTM e suas variantes, por serem comprovadas, pela literatura atual, como superiores em previsões de séries temporais.

Foram explorados teoricamente e enfrentados os desafios próprios ao processo de modelagem, como o risco de sobreajuste e a disponibilidade de dados insuficientes ou imprecisos. Para superá-los, adotamos práticas essenciais, como a compreensão do domínio do problema, seleção cuidadosa dos recursos de entrada e aplicação de métodos de regularização e validação. Além disso, exploramos a interpretabilidade dos modelos, utilizando métodos dos valores SHAP, visando evitar o tratamento dos modelos neurais como caixas pretas sempre que possível.

Ao longo das três fases de experimentos progressivos, ampliamos os horizontes dos resultados, incorporando técnicas de pré-processamento e uma maior variedade de tipos de dados. As melhorias estatísticas observadas nas diferentes etapas de pré-processamento e adição de dados pertinentes corroboram a importância e os benefícios de estratégias cuidadosas na manipulação de dados para a eficácia dos modelos de previsão. Esses resultados sugerem um potencial significativo para otimizar ainda mais os modelos e aprimorar as previsões futuras da demanda de cimento.

A análise de explicabilidade por meio dos valores SHAP confirmou a influência significativa das variáveis socioeconômicas relacionadas à construção civil e as de poder de compra da população no domínio do problema. O uso dessa ferramenta ressalta a importância crucial dessa análise, tanto quanto, que tal abordagem deve ser incorporada em qualquer atividade de predição que busque uma compreensão mais profunda do cenário de aplicação.

Este estudo fornece um panorama abrangente sobre as nuances das previsões de séries temporais de consumo de cimento, destacando a importância crítica de práticas essenciais na compreensão dos modelos. Esperamos que este trabalho seja um ponto de partida para futuras pesquisas na área de previsão por extrapolação de séries temporais, visando aprimorar ainda mais as previsões. Tais previsões são cruciais para a indústria, orientando estratégias de produção e logística, desempenhando um papel vital na tomada de decisões futuras.

# Apêndice A

## Caso referência - Benchmark

Devido ao aprofundamento deste trabalho nas dinâmicas de previsão de séries temporais, métodos e boas práticas de treinamento de modelos de aprendizado de máquina, análise minuciosa do domínio do problema e, principalmente, a vantagem de ter sido feito depois, ele teve profundas diferenças frente ao trabalho similar do ano anterior<sup>1</sup>.

Devido à impraticabilidade de usar os dados do mês presente (ou ano presente) para prever o ano futuro, foi projetado para esse trabalho um recuo de 2 anos e não somente 1 para os dados de entrada. Também, não foi utilizada uma extrapolação maior que 12 meses também devido ao fato de que, se já, é dificultoso o acesso aos dados de entrada do presente ano, a obtenção de dados de entrada do futuro é impossível sem também extrapolarmos os mesmos. Também, não foi utilizado o melhor resultado dos modelos dentro da amostra de teste externa, devido a isso ser impraticável (não podemos “olhar” para o futuro para tomarmos decisões no presente). Foi sim, utilizado a amostra de treino, que o modelo não usa para treino, mas para validação, portanto o resultado referente aos meses de 2019, seriam aqueles utilizados pela indústria na necessidade de previsão de consumo, em 2018.

Contudo, antes de iniciar as experiências relatadas nesta monografia, foi realizado testes nos mesmos moldes do trabalho anterior<sup>2</sup>, ou seja, recuo de somente um ano e extrapolação cobrindo a janela de julho de 2017 a dezembro de 2019. Isso foi feito como referência de desempenho (*benchmark*), isto é, um validador dos benefícios dos métodos e inclusões, apresentados neste trabalho, antes de mergulhar nos testes das fases apresentadas (4). Foram usados todos os métodos de pré-processamento e inclusão de dados usados na fase 3 (4.1.3), exceto a entrada de sazonalidade média. Os resultados foram os seguintes:

---

<sup>1</sup>“Previsão de consumo de cimento nos estados do Brasil usando métodos de aprendizado automático” - <https://linux.ime.usp.br/~jleite/mac0499/>

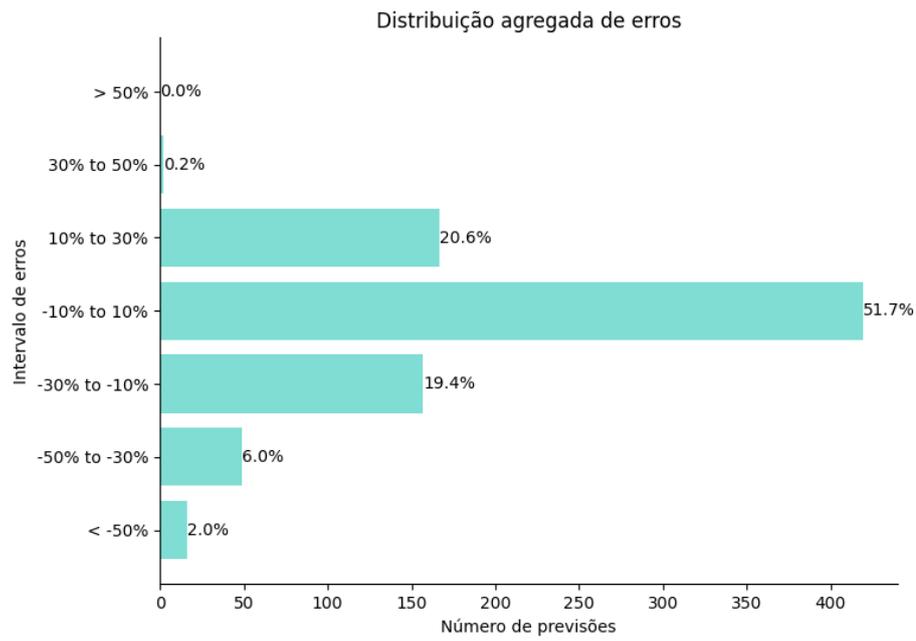
<sup>2</sup> Exceto o recuo que foi de um ano, referente ao alvo de predição, mesmo para a entradas de granularidade mensal.

Estado	RMSE (mil ton.)	MAE (mil ton.)	MAPE %	Desv.Padrão <sup>a</sup> %
Acre	2.54	2.22	17.58	10,49
Alagoas	3.53	2.68	9.29	9,08
Amapá	1.71	1.33	14.84	13,74
Amazonas	13.46	10.26	35.04	
Bahia	28.49	21.54	8.74	46,92
Ceara	21.26	16.63	12.65	8,59
Distrito Federal	6.25	5.12	10.21	12,16
Espirito Santo	11.80	8.18	9.41	7,57
Goias	23.23	18.41	9.64	11,48
Maranhão	12.44	11.15	17.08	8,05
Mato Grosso	11.90	9.91	8.99	9,63
Mato Grosso do Sul	8.92	7.38	10.76	6,97
Minas Gerais	80.49	63.15	13.30	8,97
Pará	20.03	15.81	15.79	10,78
Paraíba	8.90	7.21	11.82	15,46
Paraná	35.09	29.09	9.19	8,94
Pernambuco	19.84	14.47	9.83	6,53
Piauí	11.11	9.52	17.41	8,09
Rio Grande do Sul	17.74	13.79	6.21	12,78
Rio de Janeiro	25.65	20.66	10.18	5,26
Rio Grande do Norte	8.00	6.74	16.50	13,66
Rondônia	6.47	5.42	17.61	13,66
Roraima	1.53	1.13	18.82	35,55
Santa Catarina	34.79	29.31	10.93	6,32
São Paulo	99.32	80.48	9.62	7,81
Sergipe	5.18	4.37	12.87	7,74
Tocantins	8.48	7.43	23.12	15,73
<b>Total</b>	<b>19.56</b>	<b>15.68</b>	<b>13.61</b>	<b>12.77</b>

**Tabela A.1:** Resultados estatísticos do experimento com dados brutos.

<sup>a</sup> Desvio padrão sobre erros percentuais absolutos.

Os resultados deste teste de referência mostram os benefícios oferecidos pelas práticas apresentadas neste trabalho. As divisões dos modelos por estado, as técnicas de pré-processamento e a inclusão de novos tipos de dados contribuíram para a melhora do MAPE geral de 18%, referente ao trabalho anterior, para a redes recorrentes para aproximadamente 13,7% neste teste inicial.



**Figura A.1:** *Distribuição agregada dos erros dos dados de benchmark.*

A distribuição agregada também revelou uma melhora de assertividade, sendo que 51,7% dos erros percentuais agregados na faixa de -10% a 10%, uma melhora de aproximadamente 11,7% do esforço de pesquisa anterior.

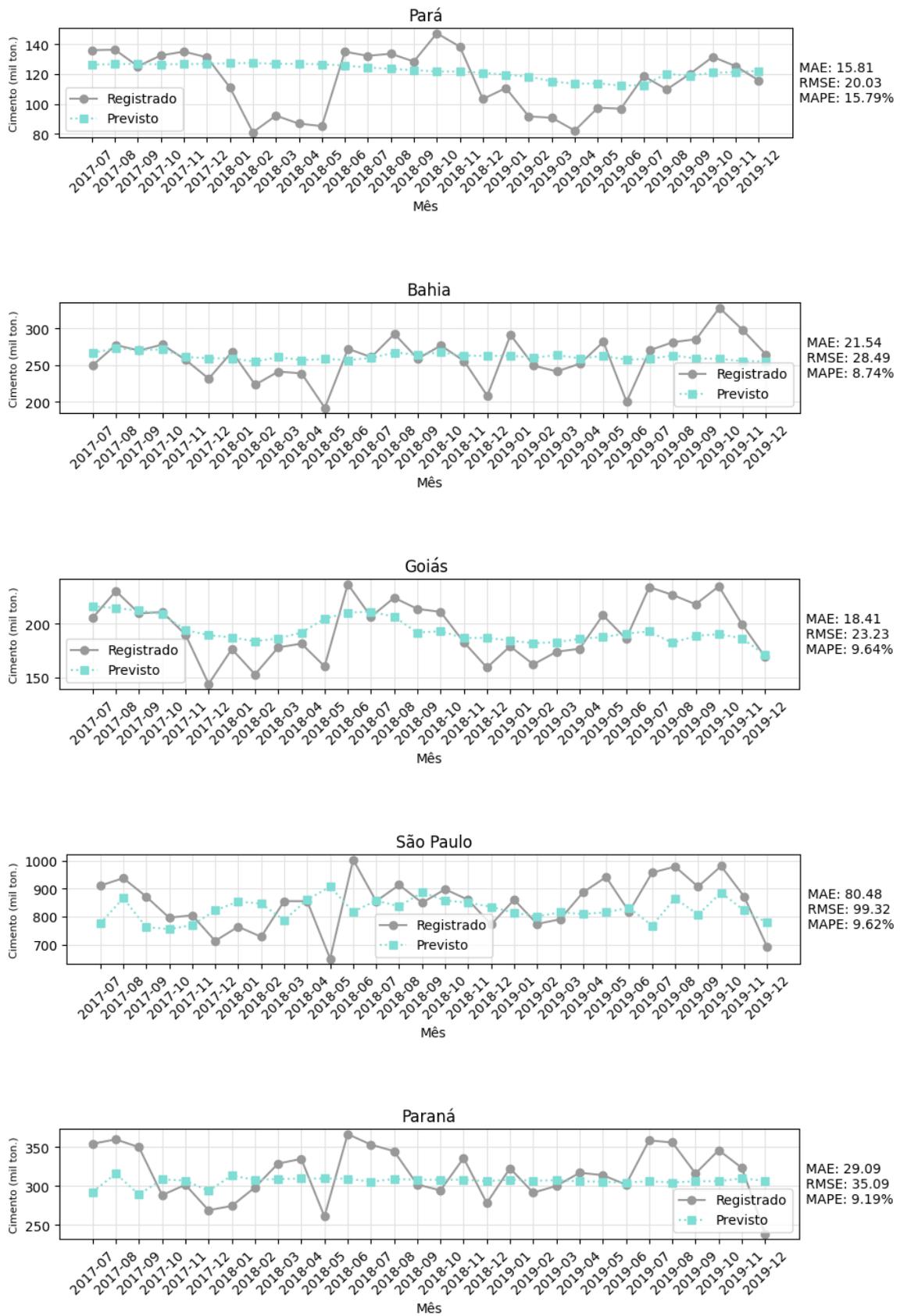


Figura A.2: Resultados das previsões de jul/2017 a dez/2019 - Dados benchmark.

## Referências

- [ABU-MOSTAFA *et al.* 2012] Yaser S. ABU-MOSTAFA, Malik MAGDON-ISMAIL e Hsuan-tien LIN. *Learning from Data: A Short Course*. AMLBook.com, 2012 (citado nas pgs. 7, 14, 15, 20, 21).
- [AGGARWAL 2018] Charu C. AGGARWAL. *Neural networks and deep learning*. Springer, 2018 (citado nas pgs. 22, 23).
- [ARAUJO 2020] Geraldo Jose Ferraresi de ARAUJO. “O coprocessamento na indústria de cimento: definição, oportunidades e vantagem competitiva”. *Revista Nacional de Gerenciamento de Cidades* 8.57 (2020), pp. 52–61 (citado na pg. 2).
- [BARRERA-ANIMAS *et al.* 2022] Ari Yair BARRERA-ANIMAS, Lukumon OYEDELE, Juan Manuel Davila DELGADO e Lukman Adewale AKANBI. “A comparative analysis of modern machine learning algorithms for time-series forecasting”. *Machine Learning with Applications* 7 (2022) (citado nas pgs. viii, 3, 15).
- [CHATFIELD 2000] Chris CHATFIELD. *Time-series forecasting*. CRC press, 2000 (citado nas pgs. 1, 6, 12, 13).
- [CHUNG *et al.* 2014] Junyoung CHUNG, Caglar GULCEHRE, KyungHyun CHO e Yoshua BENGIO. “Empirical evaluation of gated recurrent neural networks on sequence modeling”. *arXiv preprint arXiv:1412.3555* (2014) (citado nas pgs. 37, 38).
- [GOODFELLOW *et al.* 2017] Ian GOODFELLOW, Yoshua BENGIO e Aaron COURVILLE. *Deep Learning*. 22<sup>a</sup> ed. The MIT Press, 2017 (citado nas pgs. 10, 15, 23, 24).
- [HAN *et al.* 2012] Jiawei HAN, Micheline KAMBER e Jian PEI. *Data mining concepts and techniques, third edition*. Morgan Kaufmann Publishers, 2012 (citado na pg. 6).
- [HUTTER *et al.* 2019] Frank HUTTER, Lars KOTTHOFF e Joaquin VANSCHOREN. *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019 (citado na pg. 23).
- [IGHALO e ADENIYI 2020] Joshua O. IGHALO e Adewale George ADENIYI. “A perspective on environmental sustainability in the cement industry”. *Waste Disposal and Sustainable Energy* 2.3 (jan. de 2020), pp. 161–164. DOI: [10.1007/s42768-020-00043-y](https://doi.org/10.1007/s42768-020-00043-y) (citado na pg. 2).

- [*Jupyter Notebook* 2023] *Jupyter notebook*. Versão 6.5.4. 2023. URL: <https://jupyter.org> (citado na pg. 42).
- [KROLLNER *et al.* 2010] Bjoern KROLLNER, Bruce VANSTONE e Gavin FINNIE. “Financial time series forecasting with machine learning techniques: a survey” (2010), pp. 25–30 (citado na pg. 22).
- [*LibreOffice Calc* 2023] *LibreOffice Calc*. <https://www.libreoffice.org/discover/calc/>. Software de planilha eletrônica. 2023 (citado na pg. 42).
- [LUNDBERG e LEE 2017] Scott M. LUNDBERG e Su-In LEE. “A unified approach to interpreting model predictions”. *Advances in neural information processing systems* 30 (2017) (citado nas pgs. 27, 28, 54).
- [MASINI *et al.* 2023] Ricardo P. MASINI, Marcelo C. MEDEIROS e Eduardo F. MENDES. “Machine learning advances for time series forecasting”. *Journal of economic surveys* 37.1 (2023), pp. 76–111 (citado nas pgs. viii, 18, 19).
- [*Matplotlib: Visualization with Python* 2023] *Matplotlib: visualization with python*. Versão 3.8.0. 2023. URL: <https://matplotlib.org> (citado na pg. 42).
- [MULLAINATHAN e SPIESS 2017] Sendhil MULLAINATHAN e Jann SPIESS. “Machine learning: an applied econometric approach”. *Journal of Economic Perspectives* 31.2 (2017), pp. 87–106. DOI: [10.1257/jep.31.2.87](https://doi.org/10.1257/jep.31.2.87) (citado na pg. 1).
- [NBR 16697 2018] NBR 16697. *Cimento Portland - Requisitos*. 2018 (citado na pg. 2).
- [NICOLA *et al.* 2020] Maria NICOLA *et al.* “The socio-economic implications of the coronavirus pandemic (covid-19): a review”. *International journal of surgery* 78 (2020), pp. 185–193 (citado na pg. 35).
- [*NumPy: Numerical Python* 2022] *Numpy: numerical python*. Versão 1.21.3. 2022. URL: <https://numpy.org> (citado nas pgs. 9, 42).
- [*Pandas: Python Data Analysis Library* 2023] *Pandas: python data analysis library*. Versão 2.0.3. 2023. URL: <https://pandas.pydata.org> (citado na pg. 42).
- [PYLE 1999] Dorian PYLE. *Data preparation for data mining*. morgan kaufmann, 1999 (citado na pg. 12).
- [*Python Programming Language* 2023] *Python programming language*. Versão 3.9.17. 2023. URL: <https://www.python.org> (citado na pg. 42).
- [QIAO *et al.* 2019] Siyuan QIAO, Haoran WANG, Chao LIU, Wei SHEN e Alan YUILLE. “Micro-batch training with batch-channel normalization and weight standardization”. *arXiv preprint arXiv:1903.10520* (2019) (citado na pg. 11).

## REFERÊNCIAS

- [RAJKOMAR *et al.* 2019] Alvin RAJKOMAR, Jeffrey Dean M.D. e Isaac KOHANE. “Machine learning in medicine”. *New England Journal of Medicine* 380.14 (mai. de 2019), pp. 1347–1358. DOI: [10.1056/NEJMra1814259](https://doi.org/10.1056/NEJMra1814259) (citado na pg. 1).
- [SANGIORGIO e DERCOLE 2020] Matteo SANGIORGIO e Fabio DERCOLE. “Robustness of lstm neural networks for multi-step forecasting of chaotic time series”. *Chaos, Solitons & Fractals* 139.6 (out. de 2020), pp. 726–728. DOI: [10.1016/j.chaos.2020.110045](https://doi.org/10.1016/j.chaos.2020.110045) (citado nas pgs. 3, 37).
- [SHAP (SHapley Additive exPlanations) 2023] *Shap (shapley additive explanations)*. Versão 0.42.1. 2023. URL: <https://github.com/slundberg/shap> (citado nas pgs. 27, 42, 54, 61).
- [SNIC 2020] SNIC. *Relatório Anual Sindicato Nacional da Indústria do Cimento*. <http://snic.org.br/numeros-relatorio-anual.php>. Accessed: 20/11/23. 2020 (citado nas pgs. 2, 29).
- [Statsmodels: Statistical Models in Python 2023] *Statsmodels: statistical models in python*. Versão 0.14.0. 2023. URL: <https://www.statsmodels.org> (citado na pg. 42).
- [TensorFlow: An Open Source Machine Learning Framework for Everyone 2023] *Tensorflow: an open source machine learning framework for everyone*. Versão 2.13.0. 2023. URL: <https://www.tensorflow.org> (citado nas pgs. 24, 42).
- [WEERTS *et al.* 2020] Hilde JP WEERTS, Andreas C MUELLER e Joaquin VANSCHOREN. “Importance of tuning hyperparameters of machine learning algorithms”. *arXiv preprint arXiv:2007.07588* (2020) (citado na pg. 38).
- [WIKIPEDIA 2023] WIKIPEDIA. *Matriz de Vandermonde*. URL: [https://pt.wikipedia.org/wiki/Matriz\\_de\\_Vandermonde](https://pt.wikipedia.org/wiki/Matriz_de_Vandermonde) (acesso em 2023) (citado na pg. 9).
- [YU *et al.* 2019] Lu YU, Jianling QU, Feng GAO, Yanping TIAN *et al.* “A novel hierarchical algorithm for bearing fault diagnosis based on stacked lstm”. *Shock and Vibration* 2019 (2019) (citado nas pgs. viii, 37).