

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

Instituto de Ciência e Tecnologia de Sorocaba- Campus de Sorocaba

FERNANDO HATA

**ANÁLISE ESTATÍSTICA E TRATAMENTO DE SÉRIES TEMPORAIS CLIMÁTICAS
COM VARIATIONAL AUTOENCODER**

Sorocaba

2025

Fernando Hata

**ANÁLISE ESTATÍSTICA E TRATAMENTO DE SÉRIES TEMPORAIS CLIMÁTICAS
COM VARIATIONAL AUTOENCODER**

Trabalho de Conclusão de Curso, apresentado à
Universidade Estadual Paulista (UNESP), Insti-
tuto de Ciência e Tecnologia de Sorocaba, So-
rocaba, para obtenção do título de Bacharel em
Engenharia de Controle e Automação.

Orientador: Prof. Dr. Leopoldo André Dutra
Lusquino Filho

Sorocaba
2025

H361a	Hata, Fernando Análise estatística e tratamento de séries temporais climáticas com variational autoencoder / Fernando Hata. -- Sorocaba, 2025 37 p. Trabalho de conclusão de curso (Bacharelado - Engenharia de Controle e Automação) - Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia, Sorocaba Orientador: Leopoldo André Dutra Lusquino Filho 1. Aprendizado do computador. 2. Estatística.. 3. Clima. 4. Python (Linguagem de programação de computador). 5. Modelos matemáticos. I. Título.
-------	--

FERNANDO HATA

**ANÁLISE ESTATÍSTICA E TRATAMENTO DE SÉRIES TEMPORAIS CLIMÁTICAS
COM VARIATIONAL AUTOENCODER**

Trabalho de Conclusão de Curso apresentado à Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia de Sorocaba, Sorocaba, para obtenção do título de Bacharel em Engenharia de Controle e Automação.

Data de defesa: 05/11/2025

BANCA EXAMINADORA

Prof. Dr. Leopoldo André Dutra Lusquino Filho
Instituto de Ciência e Tecnologia de Sorocaba – Campus de Sorocaba

Profº Dr. Eduardo Verri Liberado
Instituto de Ciência e Tecnologia de Sorocaba – Campus de Sorocaba

BSC Sidney Alves de Outeiro
Universidade Federal do Rio de Janeiro

Dedico este trabalho à família Hata.

AGRADECIMENTOS

A Deus, pela força e resiliência concedidas.

À minha família, pela fé e persistência.

À república HM, por me acompanhar.

À Julia Alves da Silva, pelos itens anteriores e todo o restante.

“A matemática é o alfabeto com o qual Deus escreveu o universo.”
Galileu Galilei

RESUMO

Este trabalho apresenta o desenvolvimento de um sistema interativo voltado à coleta, validação, reconstrução e visualização de dados meteorológicos, com ênfase na análise estatística descritiva e na aplicação de técnicas de aprendizado de máquina, especificamente o Variational Autoencoder (VAE). A solução proposta integra, em uma única interface, múltiplas etapas do fluxo de análise, incluindo a extração automatizada de informações de estações climáticas, o processamento baseado em critérios de qualidade e a geração de visualizações dinâmicas. O sistema foi implementado utilizando a biblioteca Streamlit, permitindo o acesso a funcionalidades avançadas de forma intuitiva e acessível a usuários com diferentes níveis de conhecimento em programação. A aplicação do VAE possibilitou a reconstrução de valores ausentes, contribuindo para a consistência e completude das séries temporais analisadas, além de fornecer meios para investigar a estrutura latente dos dados. O referencial teórico que sustenta este trabalho abrange conceitos fundamentais da estatística descritiva, tais como medidas de tendência central, dispersão, assimetria, curtose e autocorrelação, bem como técnicas modernas de reconstrução de dados. Os experimentos realizados evidenciaram a eficiência e a reprodutibilidade da abordagem, demonstrando que o sistema atende às demandas de pesquisadores e instituições que necessitam de ferramentas confiáveis e organizadas para a análise de séries temporais climáticas. Conclui-se que a proposta apresentada oferece contribuições relevantes tanto em termos de automação e qualidade analítica quanto em aderência a boas práticas de Engenharia de Software, destacando-se como um exemplo de aplicação acadêmica de integração entre ciência de dados e engenharia de sistemas.

Palavras-Chave: séries temporais; dados climáticos; autoencoders variacionais; análise estatística; streamlit.

ABSTRACT

This work presents the development of an interactive system designed for the collection, validation, reconstruction, and visualization of meteorological data, with emphasis on descriptive statistical analysis and the application of machine learning techniques, specifically the Variational Autoencoder (VAE). The proposed solution integrates multiple stages of the analytical workflow into a single interface, including automated extraction of information from climate stations, quality-based data processing, and dynamic visualizations. The system was implemented using the Streamlit library, enabling access to advanced functionalities through an intuitive and user-friendly environment for researchers with different levels of programming expertise. The application of the VAE enabled the reconstruction of missing values, contributing to the consistency and completeness of the analyzed time series, in addition to providing a way to investigate the latent structure of the data. The theoretical framework supporting this work encompasses fundamental concepts of descriptive statistics, such as measures of central tendency, dispersion, skewness, kurtosis, and autocorrelation, as well as modern techniques for data reconstruction. The experiments carried out demonstrated the efficiency and reproducibility of the approach, showing that the system meets the demands of researchers and institutions that require reliable and organized tools for the analysis of climate time series. It is concluded that the proposed system provides relevant contributions both in terms of automation and analytical quality and in adherence to software engineering best practices, standing out as an academic example of the integration between data science and systems engineering.

Keywords: time series; climate data; variational autoencoders; statistical analysis; streamlit.

LISTA DE ILUSTRAÇÕES

Figura 1	Histograma com curva de densidade (KDE) da variável temperatura. . . .	18
Figura 2	Boxplot da variável temperatura.	19
Figura 3	Série temporal da temperatura.	19
Figura 4	Correlogramas ACF e PACF para a variável temperatura.	19
Figura 5	Mapa de calor das correlações entre variáveis climáticas.	20
Figura 6	Arquitetura conceitual do VAE.	21
Figura 7	Fluxo geral da aplicação proposta, com etapas de coleta, validação, reconstrução e visualização.	24
Figura 8	Interface de coleta de dados meteorológicos conectada à base da NOAA. .	25
Figura 9	Interface de seleção de estação, período de análise e threshold de dados válidos.	26
Figura 10	Pré-visualização dos dados após o filtro de validade, com resumo estatístico.	26
Figura 11	Interface de configuração dos parâmetros do modelo VAE.	27
Figura 12	Exemplo de visualização de dados originais e reconstruídos após execução do VAE.	28
Figura 13	Série temporal da variável TMIN.	29
Figura 14	Histograma com curva de densidade (KDE) para a variável TMIN.	29
Figura 15	Boxplot da variável TMIN com destaque para outliers.	29
Figura 16	Correlogramas ACF e PACF da variável TMIN.	30
Figura 17	Mapa de calor da matriz de correlação envolvendo a variável TMIN. . . .	30
Figura 18	Painel interativo com estatísticas descritivas e visualizações individuais por variável numérica.	31
Figura 19	Comparação de séries temporais para a variável TMIN: RawData (azul) e VAE (vermelho tracejado).	31
Figura 20	Alinhamento entre séries RawData e VAE da variável TMIN utilizando a distância DTW.	32

LISTA DE SIGLAS

ABNT Associação Brasileira de Normas Técnicas

BACEN Banco Central do Brasil

DTW Dynamic Time Warping

NOAA National Oceanic and Atmospheric Administration

VAE Variational Autoencoder

LISTA DE SÍMBOLOS

A	Amplitude total
D_{KL}	Divergência de Kullback-Leibler
E	Valor esperado (Esperança matemática)
$freq(x_i)$	Frequência de ocorrência do valor x_i
I	Matriz identidade
MAE	Erro Absoluto Médio
n	Número total de observações na amostra
$N(0, I)$	Distribuição Normal Padrão
$p(x)$	Evidência marginal dos dados
$p(z)$	Distribuição a priori das variáveis latentes
$p_\theta(x z)$	Modelo probabilístico do <i>Decoder</i> (verossimilhança)
$q_\phi(z x)$	Modelo probabilístico do <i>Encoder</i> (aproximação variacional)
r	Coefficiente de correlação linear de Pearson
R^2	Coefficiente de determinação
$RMSE$	Raiz do Erro Quadrático Médio
s	Desvio-padrão amostral
s^2	Variância amostral
x	Vetor de dados de entrada ou variável observada
$\bar{x}$	Média aritmética da amostra
x_i	i -ésimo valor da variável observada
$x_{(i)}$	Valores ordenados da amostra (estatísticas de ordem)
x_{max}	Valor máximo da amostra
x_{min}	Valor mínimo da amostra
$\bar{y}$	Média aritmética da segunda variável (para correlação)

y_i	i -ésimo valor da segunda variável (para correlação)
z	Vetor de variáveis latentes
ϵ	Variável de ruído auxiliar
θ	Parâmetros treináveis da rede neural do <i>Decoder</i>
μ	Média da distribuição latente
σ	Desvio-padrão da distribuição latente
ϕ	Parâmetros treináveis da rede neural do <i>Encoder</i>

SUMÁRIO

1	INTRODUÇÃO	14
2	REFERENCIAL TEÓRICO	16
2.1	Análise Estatística Descritiva	16
2.1.1	Medidas de Tendência Central	16
2.1.1.1	Média Aritmética	16
2.1.1.2	Mediana	16
2.1.1.3	Moda	16
2.1.2	Medidas de Dispersão	17
2.1.2.1	Variância	17
2.1.2.2	Desvio-Padrão	17
2.1.2.3	Amplitude Total	17
2.1.3	Medidas de Forma	17
2.1.3.1	Assimetria (Skewness)	17
2.1.3.2	Curtose	17
2.1.4	Correlação Linear	18
2.1.4.1	Coeficiente de Correlação de Pearson	18
2.2	Visualizações Gráficas	18
2.2.1	Histogramas com Curva de Densidade (KDE)	18
2.2.2	Boxplots Interativos	18
2.2.3	Séries Temporais	18
2.2.4	Correlogramas ACF e PACF	19
2.2.5	Mapa de Calor Interativo de Correlação	19
2.3	Variational Autoencoder (VAE)	20
2.3.1	Arquitetura do VAE	20
2.3.2	Formulação Matemática	20
2.3.3	Aplicação em Séries Temporais	21
3	TRABALHOS CORRELATOS	22
3.1	Silva et al. (2019)	22
3.2	Pereira e Oliveira (2020)	22
3.3	Santos et al. (2021)	22
3.4	Zhou et al. (2022)	22
3.5	Kaur e Bansal (2023)	23
4	DESENVOLVIMENTO / METODOLOGIA	24

4.1	Abordagem Proposta	24
4.1.1	Coleta de Dados	24
4.1.2	Extração de Dados Válidos	25
4.1.3	Processamento VAE	27
4.1.4	Processamento Gráfico	28
4.1.4.1	Séries Temporais Interativas	28
4.1.4.2	Histograma com Curva de Densidade (KDE)	29
4.1.4.3	Boxplots Interativos	29
4.1.4.4	Correlogramas ACF e PACF	29
4.1.4.5	Mapa de Calor Interativo de Correlação	30
5	DISCUSSÃO DOS RESULTADOS	33
6	CONCLUSÃO	36
	REFERÊNCIAS	37

1 INTRODUÇÃO

O aquecimento global e as mudanças climáticas representam um dos maiores desafios científicos e sociais do século XXI. Eventos extremos, como ondas de calor, secas prolongadas e tempestades intensas, têm se tornado cada vez mais frequentes e severos, afetando diretamente a disponibilidade de recursos hídricos, a segurança alimentar, a biodiversidade e a qualidade de vida das populações (Intergovernmental Panel on Climate Change, 2021).

Nesse contexto, a análise de séries temporais meteorológicas — volumosas, ruidosas e frequentemente incompletas — torna-se indispensável para detectar padrões, identificar anomalias e apoiar políticas de mitigação e adaptação (Storch; Zwiers, 2006).

Diversas ferramentas especializadas para o tratamento de dados climáticos já estão disponíveis, como o CLIMSOFT, o RClimDex, o Panoply e o Climate Data Operators (CDO). Apesar de úteis, essas soluções geralmente são voltadas a tarefas específicas — como visualização isolada, análise de extremos ou manipulação via linha de comando — e carecem de uma interface integrada, acessível e interativa para análise completa de dados (Max Planck Institute for Meteorology, 2020; World Meteorological Organization, 2022; Zhang; Yang, 2009).

Para superar essas limitações, este trabalho propõe o desenvolvimento de um sistema interativo *end-to-end*, construído em Python com as bibliotecas *Streamlit* (Streamlit Inc., 2023), *Plotly* (Plotly Technologies Inc., 2015) e *Statsmodels* (Seabold; Perktold, 2010), que unifica todo o fluxo de análise de dados climáticos (McKinney, 2010; Hunter, 2007).

A plataforma facilita:

- Coleta automática de dados diretamente dos repositórios oficiais da National Oceanic and Atmospheric Administration (NOAA);
- Filtragem e validação de dados com base em critérios de ausência máxima de valores;
- Reconstrução de séries temporais por meio de Variational Autoencoder (VAE) (*Variational Autoencoder*) (Kingma; Welling, 2014);
- Geração de visualizações interativas — histogramas, boxplots, ACF/PACF e mapas de calor;
- Cálculo de métricas como distância Dynamic Time Warping (DTW) (*Dynamic Time Warping*) (Berndt; Clifford, 1994).

A interface foi concebida para atender pesquisadores, profissionais e estudantes interessados em análise climática, oferecendo um ambiente intuitivo, altamente parametrizável e orientado à reprodutibilidade.

Esta dissertação está organizada em cinco capítulos. O Capítulo 2 apresenta o referencial teórico. O Capítulo 3 descreve a abordagem proposta e os módulos do sistema. O Capítulo 4

discute vantagens, limitações e comparações com soluções existentes. Por fim, o Capítulo 5 sintetiza as conclusões e propõe extensões futuras.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta a fundamentação teórica que embasa as técnicas estatísticas e de aprendizagem profunda aplicadas ao processamento e análise de séries temporais meteorológicas.

2.1 ANÁLISE ESTATÍSTICA DESCRITIVA

A análise estatística descritiva é um conjunto de técnicas voltadas à sumarização, exploração e compreensão inicial de dados. No contexto de séries temporais meteorológicas, essas técnicas possibilitam a identificação de padrões, tendências, valores atípicos e possíveis relações entre variáveis. A seguir, apresentam-se as principais medidas utilizadas, suas definições matemáticas e representações visuais associadas.

2.1.1 Medidas de Tendência Central

As medidas de tendência central representam valores em torno dos quais os dados tendem a se concentrar. As três medidas mais utilizadas são: média, mediana e moda.

2.1.1.1 Média Aritmética

A média aritmética é obtida pela soma de todos os valores dividida pelo número total de observações, conforme definido na Equação (1).

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

2.1.1.2 Mediana

A mediana é o valor central de um conjunto de dados ordenado. Para conjuntos com número ímpar de elementos, é o valor na posição central; para número par, é a média dos dois valores centrais, como ilustrado na Equação (2).

$$\text{Mediana} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ for ímpar} \\ \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right), & \text{se } n \text{ for par} \end{cases} \quad (2)$$

2.1.1.3 Moda

A moda é o valor que ocorre com maior frequência no conjunto de dados, sendo definida pela Equação (3).

$$\text{Moda} = \arg \max_{x_i} \text{freq}(x_i) \quad (3)$$

2.1.2 Medidas de Dispersão

As medidas de dispersão expressam o grau de variabilidade dos dados em relação à média. As principais são: variância, desvio-padrão e amplitude total.

2.1.2.1 Variância

A variância representa o desvio quadrático médio em relação à média e é dada pela Equação (4).

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4)$$

2.1.2.2 Desvio-Padrão

O desvio-padrão corresponde à raiz quadrada da variância e é representado pela Equação (5).

$$s = \sqrt{s^2} \quad (5)$$

2.1.2.3 Amplitude Total

A amplitude é a diferença entre o maior e o menor valor da amostra, conforme Equação (6).

$$A = x_{\text{máx}} - x_{\text{mín}} \quad (6)$$

2.1.3 Medidas de Forma

Essas medidas descrevem a forma da distribuição dos dados em relação à média, identificando assimetrias e o achatamento da curva.

2.1.3.1 Assimetria (Skewness)

A assimetria quantifica o grau de simetria da distribuição em torno da média. Sua fórmula é apresentada na Equação (7).

$$\text{Assimetria} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad (7)$$

2.1.3.2 Curtose

A curtose avalia o grau de concentração dos dados nas extremidades (caudas) da distribuição, como definido na Equação (8).

$$\text{Curtose} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 \quad (8)$$

2.1.4 Correlação Linear

2.1.4.1 Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson mede a relação linear entre duas variáveis, e sua fórmula é apresentada na Equação (9).

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (9)$$

2.2 VISUALIZAÇÕES GRÁFICAS

A visualização de dados potencializa a interpretação intuitiva dos resultados estatísticos. As ferramentas adotadas no sistema incluem diferentes representações visuais que auxiliam na detecção de padrões e anomalias em séries temporais. A seguir, cada técnica é descrita, acompanhada de uma ilustração gerada pelo sistema.

2.2.1 Histogramas com Curva de Densidade (KDE)

Permitem observar a distribuição de frequência de uma variável contínua, com sobreposição de uma estimativa suave da densidade de probabilidade. Essa combinação ajuda a identificar a forma da distribuição e presença de valores extremos.

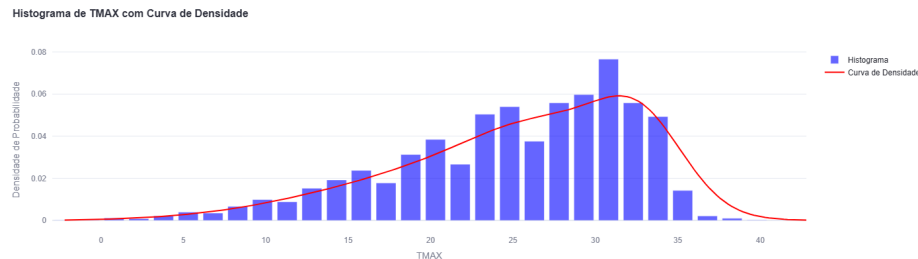


Figura 1 – Histograma com curva de densidade (KDE) da variável temperatura.

2.2.2 Boxplots Interativos

Representam a distribuição de uma variável com base em seus quartis, evidenciando a mediana, os limites inferior e superior e possíveis outliers. São úteis para comparar a dispersão entre variáveis ou períodos distintos.

2.2.3 Séries Temporais

Exibem a variação de uma variável ao longo do tempo. Essa visualização é essencial para investigar tendências, sazonalidade e descontinuidades nos dados.

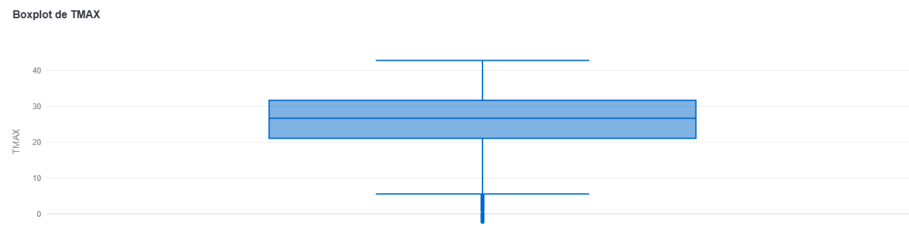


Figura 2 – Boxplot da variável temperatura.

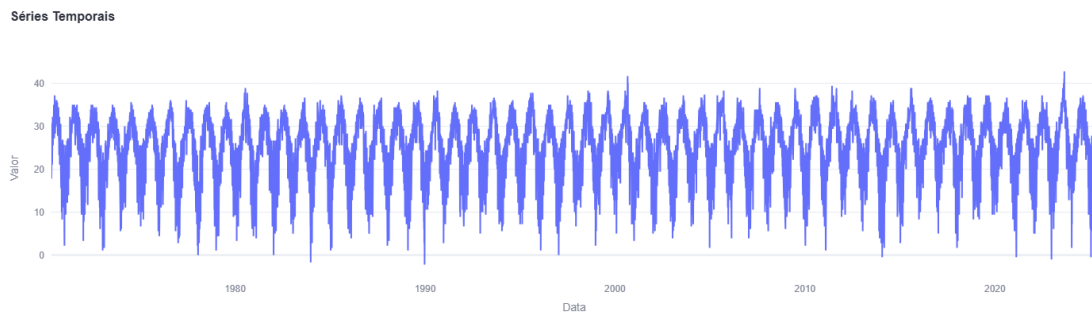


Figura 3 – Série temporal da temperatura.

2.2.4 Correlogramas ACF e PACF

O correlograma de autocorrelação (ACF) e o de autocorrelação parcial (PACF) indicam dependências entre observações em diferentes defasagens. São amplamente utilizados na modelagem de séries temporais (como ARIMA).

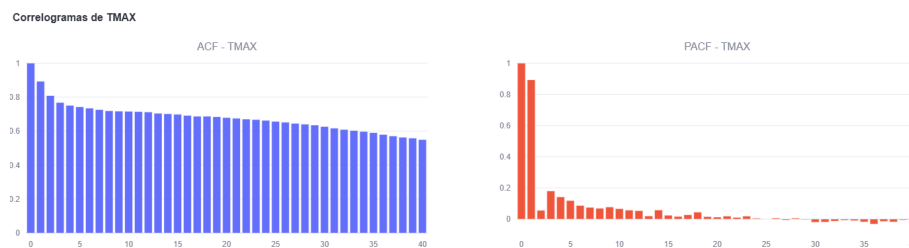


Figura 4 – Correlogramas ACF e PACF para a variável temperatura.

2.2.5 Mapa de Calor Interativo de Correlação

Exibe uma matriz de correlação entre variáveis. Cores mais intensas representam correlações mais fortes (positivas ou negativas), facilitando a identificação de relações lineares entre variáveis climáticas.

Essas visualizações são particularmente úteis na identificação de tendências sazonais, anomalias, valores extremos e padrões latentes em séries temporais.



Figura 5 – Mapa de calor das correlações entre variáveis climáticas.

2.3 VARIATIONAL AUTOENCODER (VAE)

Os *Variational Autoencoders* (VAEs) são modelos generativos probabilísticos baseados em redes neurais profundas, propostos por Kingma e Welling (Kingma; Welling, 2014). Seu objetivo é aprender representações latentes eficientes e contínuas que permitam gerar novas amostras com propriedades similares às do conjunto de dados original.

Ao contrário de autoencoders tradicionais, que comprimem a entrada em um ponto fixo do espaço latente, os VAEs aprendem uma distribuição sobre o espaço latente, geralmente assumida como Gaussiana. Essa abordagem permite gerar amostras com coerência estrutural e estimar incertezas.

2.3.1 Arquitetura do VAE

A arquitetura do VAE consiste em três componentes:

- **Encoder (Reconhecimento):** mapeia uma entrada x para os parâmetros (μ, σ^2) de uma distribuição normal latente $q_\phi(z|x)$.
- **Amostragem Latente:** utiliza a técnica de reparametrização: $z = \mu + \sigma \cdot \epsilon$, com $\epsilon \sim \mathcal{N}(0, I)$.
- **Decoder (Geração):** reconstrói x a partir da amostra z via $p_\theta(x|z)$.

2.3.2 Formulação Matemática

O VAE maximiza a evidência marginal dos dados $p(x)$ por meio de uma aproximação variacional:

$$\log p(x) \geq E_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{\text{KL}}[q_\phi(z|x)||p(z)] \quad (10)$$

Essa desigualdade define a função objetivo do VAE, conhecida como **ELBO** (Evidence Lower Bound), composta por dois termos:

- $E_{q_\phi(z|x)}[\log p_\theta(x|z)]$: verossimilhança esperada da reconstrução;

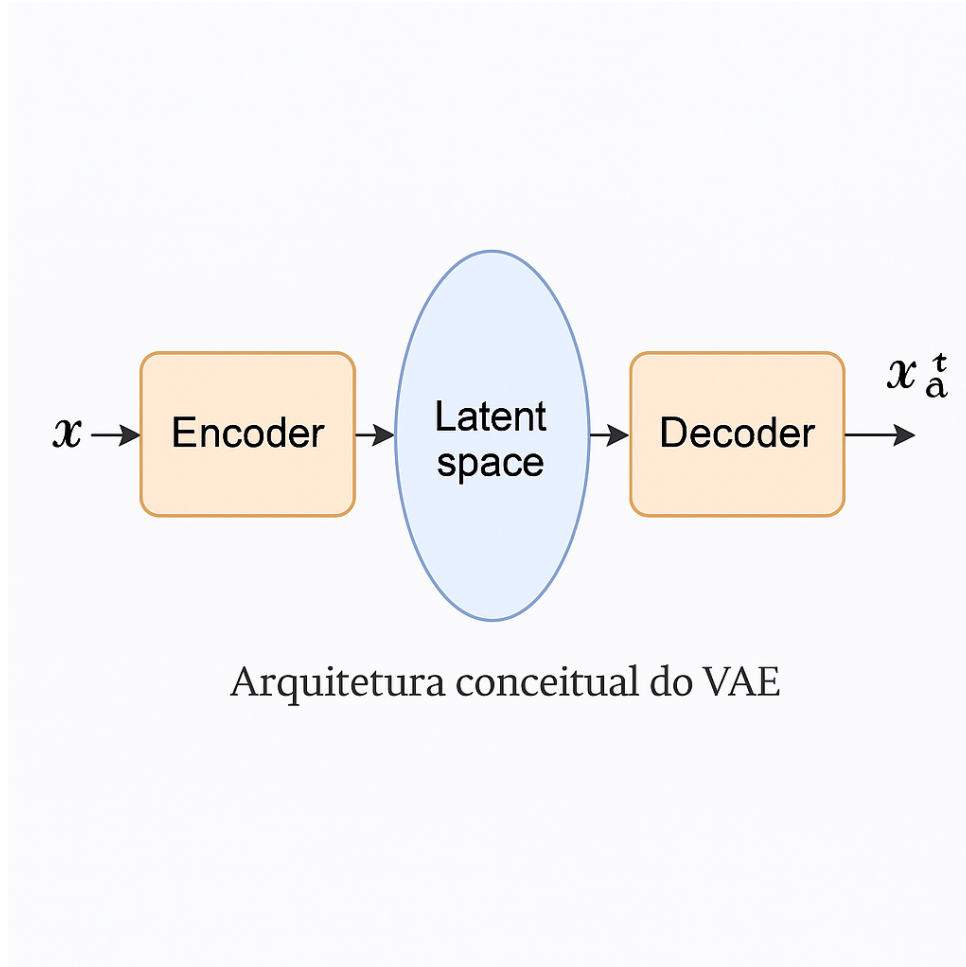


Figura 6 – Arquitetura conceitual do VAE.

- $D_{\text{KL}}[q_{\phi}(z|x)||p(z)]$: divergência de Kullback–Leibler entre a distribuição latente inferida e a *prior*.

2.3.3 Aplicação em Séries Temporais

A utilização de VAEs em séries temporais visa reconstruir pontos ausentes ou ruidosos com base na estrutura latente aprendida a partir do histórico da série. Isso é particularmente relevante em dados meteorológicos, que frequentemente apresentam falhas de medição.

Neste trabalho, o VAE foi adaptado para lidar com entradas sequenciais, mantendo a suposição de independência temporal local em janelas fixas. A arquitetura é composta por camadas densas e os hiperparâmetros são ajustáveis via interface. A abordagem baseia-se nos estudos de (Kingma; Welling, 2014) e explora aplicações já documentadas na literatura nacional, como a de Pereira e Oliveira (Pereira; Oliveira, 2020) na modelagem de precipitação.

3 TRABALHOS CORRELATOS

Diversos autores têm explorado o uso de técnicas estatísticas e de aprendizado de máquina na análise de dados climáticos, buscando aprimorar tarefas como detecção de anomalias, reconstrução de séries, previsão de eventos extremos e visualização exploratória.

3.1 SILVA ET AL. (2019)

Aplicaram autoencoders clássicos à reconstrução de séries temporais de temperatura obtidas de estações meteorológicas brasileiras. O foco foi detectar anomalias sazonais e ruídos instrumentais. A técnica obteve resultados superiores aos métodos tradicionais de imputação (como média móvel), principalmente na manutenção da estrutura periódica dos dados e na redução do erro quadrático médio (RMSE) em até 35%.

3.2 PEREIRA E OLIVEIRA (2020)

Utilizaram Variational Autoencoders (VAEs) para previsão de precipitação diária com base em séries temporais históricas da região Sul do Brasil. Compararam o desempenho com LSTM (Long Short-Term Memory) e modelos SARIMA. Os VAEs mostraram-se superiores em termos de generalização e reconstrução de padrões com falhas, alcançando aumento de 12% no coeficiente de determinação (R^2) e menor erro absoluto médio (MAE).

3.3 SANTOS ET AL. (2021)

Combinaram mapas de calor de correlação e técnicas de agrupamento hierárquico para investigar padrões regionais em séries de temperatura, umidade e radiação solar no estado de Minas Gerais. A abordagem permitiu identificar grupos de estações com comportamento semelhante e relações espaço-temporais significativas. A principal contribuição foi uma metodologia exploratória para delimitação de zonas climáticas homogêneas.

3.4 ZHOU ET AL. (2022)

Propuseram um modelo híbrido baseado em CNN-LSTM para previsão de séries meteorológicas multivariadas em datasets públicos como WeatherBench. A técnica foi avaliada com dados de temperatura, pressão e vento, e superou modelos ARIMA e LSTM simples, obtendo até 18% de melhoria no RMSE em horizonte de 5 dias (Zhou; Chen; Wang, 2022).

3.5 KAUR E BANSAL (2023)

Implementaram um framework baseado em autoencoders profundos para detectar e corrigir falhas em sensores climáticos em tempo real. Os testes foram conduzidos com dados da rede NOAA, e os resultados mostraram redução de 22% no tempo médio de resposta em comparação com métodos manuais (Kaur; Bansal, 2023).

Embora esses estudos contribuam com avanços relevantes em tarefas específicas, o presente trabalho inova ao integrar, em uma única plataforma interativa, todo o pipeline de análise — incluindo coleta automática, validação, reconstrução com VAE, análise estatística descritiva e visualização gráfica dinâmica. Essa abordagem unificada visa oferecer maior reprodutibilidade, facilidade de uso e capacidade analítica para estudos exploratórios em dados climáticos incompletos.

4 DESENVOLVIMENTO / METODOLOGIA

4.1 ABORDAGEM PROPOSTA

Neste capítulo, apresentamos a aplicação prática dos conceitos discutidos no referencial teórico, com foco na implementação do sistema interativo desenvolvido para análise e visualização de dados meteorológicos. A proposta contempla um conjunto de interfaces funcionais desenvolvidas com Streamlit, que integram técnicas estatísticas descritivas, reconstrução via VAE (Variational Autoencoder) e mecanismos interativos de comparação entre dados originais e reconstruídos.

As seções seguintes descrevem, individualmente, cada módulo funcional da aplicação, ressaltando suas características, objetivos e a fundamentação teórica aplicada — conforme discutido no capítulo 2. A proposta também enfatiza a usabilidade da ferramenta, permitindo ao usuário escolher a estação meteorológica, o período de análise, o tipo de dado (RawData ou VAE), bem como visualizar, salvar e comparar os resultados de maneira intuitiva e eficaz.

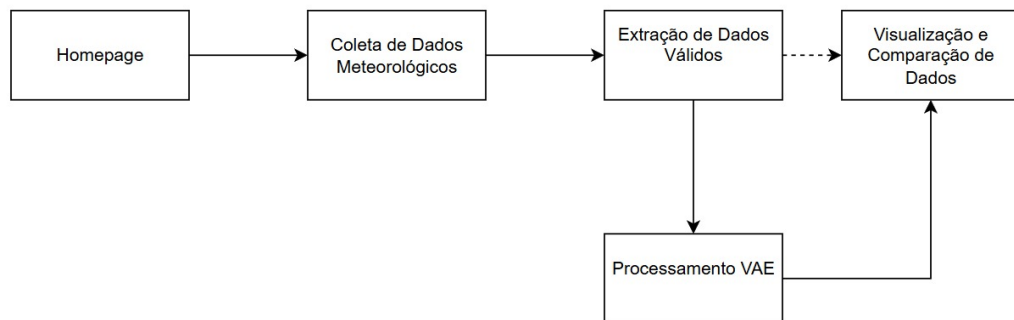


Figura 7 – Fluxo geral da aplicação proposta, com etapas de coleta, validação, reconstrução e visualização.

4.1.1 Coleta de Dados

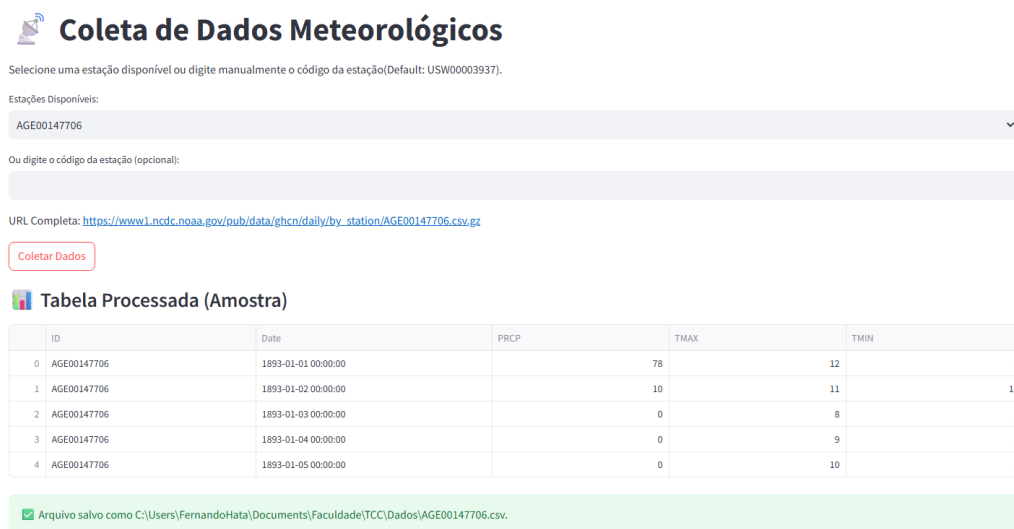
Esta etapa é responsável pela coleta ativa de dados meteorológicos históricos diretamente da base oficial da NOAA (National Oceanic and Atmospheric Administration). A aplicação se conecta ao diretório remoto da NOAA e recupera a lista de estações meteorológicas disponíveis, permitindo que o usuário selecione uma delas por meio de um menu interativo ou digite manualmente o código da estação desejada.

Após a seleção, o sistema constrói automaticamente a URL de acesso ao arquivo `.csv.gz` correspondente à estação escolhida. Esses arquivos contêm dados diários codificados em formato padronizado. O sistema realiza o download, descompactação e transformação dos dados em uma estrutura tabular adequada para análise. Esse processo inclui:

- Conversão da coluna de datas para o formato `datetime`;

- Reestruturação da tabela para formato *pivotado*, onde cada variável climática ocupa uma coluna distinta;
- Conversão automática das variáveis de temperatura de décimos de grau Celsius para graus Celsius.

Após o processamento, uma amostra dos dados é exibida na interface. O arquivo final é salvo localmente em um diretório pré-definido, pronto para uso nas etapas posteriores do sistema.



Coleta de Dados Meteorológicos

Selecione uma estação disponível ou digite manualmente o código da estação(Default: USW00003937).

Estações Disponíveis:

AGE00147706

Ou digite o código da estação (opcional):

URL Completa: https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_station/AGE00147706.csv.gz

Coletar Dados

Tabela Processada (Amostra)

	ID	Date	PRCP	TMAX	TMIN	
0	AGE00147706	1893-01-01 00:00:00		78	12	9
1	AGE00147706	1893-01-02 00:00:00		10	11	10
2	AGE00147706	1893-01-03 00:00:00		0	8	5
3	AGE00147706	1893-01-04 00:00:00		0	9	5
4	AGE00147706	1893-01-05 00:00:00		0	10	6

Arquivo salvo como C:\Users\FernandoHata\Documents\Faculdade\TCC\Dados\AGE00147706.csv.

Figura 8 – Interface de coleta de dados meteorológicos conectada à base da NOAA.

4.1.2 Extração de Dados Válidos

Após a coleta inicial, os dados brutos são submetidos a uma etapa de filtragem e validação, cujo objetivo é assegurar a consistência e a qualidade da amostra utilizada nas análises estatísticas.

Primeiramente, os dados são convertidos para o tipo de data (`datetime`) e ordenados cronologicamente. Em seguida, aplica-se um filtro de intervalo de datas definido pelo usuário, o que permite restringir a análise a um período específico de interesse. Após essa etapa, o sistema calcula a proporção de valores ausentes em cada coluna e valida apenas aquelas cujo percentual de faltantes é inferior a um limiar pré-estabelecido.

As colunas que não satisfazem esse critério são consideradas inválidas e desconsideradas nas análises subsequentes. O sistema também gera um resumo textual, indicando a quantidade de valores válidos e ausentes para cada variável, bem como a taxa de perda nas colunas descartadas.

A partir das variáveis validadas, é criado um novo conjunto de dados contendo apenas registros completos (sem valores nulos). Este conjunto final é salvo automaticamente com o sufixo `CLEAN_YYYY-MM-DD_YYYY-MM-DD`, indicando o intervalo temporal ao qual os dados se referem.

 **Extração de Dados Válidos** 

Selecione uma estação:

USW00003937

Período de Análise

Data Inicial:

1980/05/02

Data Final:

2025/05/17

Data mínima encontrada no arquivo: 1942-05-02 00:00:00

Data máxima encontrada no arquivo: 2025-05-17 00:00:00

Threshold para valores ausentes (0 a 1)

0,10

Extrair Dados Válidos

Arquivo extraído: C:\Users\FernandoHata\Documents\Faculdade\TCC\Plotagem\USW00003937\USW00003937_CLEAN_1980-05-02_2025-05-17.csv

Figura 9 – Interface de seleção de estação, período de análise e threshold de dados válidos.

	ID	Date	AWND	PRCP	TMAX	TMIN
14065	USW00003937	1984-01-01 00:00:00	34	0	15	-1.7
14066	USW00003937	1984-01-02 00:00:00	28	0	15.6	2.8
14067	USW00003937	1984-01-03 00:00:00	25	0	16.7	5
14068	USW00003937	1984-01-04 00:00:00	36	0	18.9	2.8
14069	USW00003937	1984-01-05 00:00:00	32	0	20.6	7.2

Coluna: ID - Válidos: 16452 / Nulos: 0

Coluna: Date - Válidos: 16452 / Nulos: 0

Coluna: AWND - Válidos: 15066 / Nulos: 1386

Coluna: PRCP - Válidos: 16451 / Nulos: 1

Coluna: TMAX - Válidos: 16451 / Nulos: 1

Coluna: TMIN - Válidos: 16451 / Nulos: 1

Colunas ignoradas:

ACMC: 1.00

ACMH: 0.65

ACSC: 1.00

ACSH: 0.65

ADPT: 0.59

ASLP: 0.59

ASTP: 0.59

AWBT: 0.59

FMTM: 0.38

Figura 10 – Pré-visualização dos dados após o filtro de validade, com resumo estatístico.

Essa etapa de preparação garante que as análises estatísticas e as visualizações subsequentes sejam aplicadas somente a dados confiáveis, respeitando padrões de completude e consistência necessários para interpretações robustas.

4.1.3 Processamento VAE

Após a validação e visualização dos dados, o sistema permite a aplicação do modelo VAE (Variational Autoencoder) sobre os dados meteorológicos selecionados. Esta etapa é realizada em uma interface específica, onde o usuário pode configurar os hiperparâmetros do modelo antes de iniciar o processamento.

Dentre os parâmetros configuráveis estão:

- Número de épocas de treinamento;
- Tamanho do *batch*;
- Tamanho do espaço latente;
- Taxa de aprendizado.

O VAE é treinado a partir dos dados validados (*RawData*), considerando apenas colunas numéricas e completas. Após o treinamento, o modelo reconstrói os dados de entrada e salva um novo arquivo contendo os valores gerados. Esse novo conjunto, denominado *VAE*, reflete a estrutura latente aprendida a partir dos dados originais e pode ser utilizado em análises comparativas.

O sistema realiza automaticamente o salvamento dos dados reconstruídos em um diretório organizado por estação, período e tipo de dado. É gerado também um arquivo com os erros de reconstrução, permitindo ao usuário avaliar a performance do modelo e identificar pontos com alto erro — o que pode indicar ruídos ou outliers relevantes.

The image shows a web interface for configuring a Variational Autoencoder (VAE) model. It includes two dropdown menus for station and file selection, followed by a section titled 'Parâmetros do VAE' with four input fields for latent dimension, epochs, batch size, and learning rate, each with minus and plus buttons for adjustment. An 'Executar VAE' button is at the bottom.

Seleccione a estação:

USW00003937

Seleccione o arquivo:

USW00003937_CLEAN_1970-05-02_2025-05-17.csv

Parâmetros do VAE

Dimensão Latente

10

Épocas

50

Tamanho do Lote

32

Taxa de Aprendizado

0,001

Executar VAE

Figura 11 – Interface de configuração dos parâmetros do modelo VAE.

Deploy

Visualizando os primeiros dados:

	ID	Date	PRCP	TMAX	TMIN
0	USW00003937	1970-05-02		226	17.8
1	USW00003937	1970-05-03		20	23.3
2	USW00003937	1970-05-04		0	26.7
3	USW00003937	1970-05-05		0	28.3
4	USW00003937	1970-05-06		0	30

Tipos de dados:

	0
ID	object
Date	object
PRCP	float64
TMAX	float64
TMIN	float64

VAE concluído! Arquivo salvo: C:\Users\FernandoHata\Documents\Faculdade\TCC\Plotagem\USW00003937\USW00003937_WRITE_1970-05-02_2025-05-17.csv

	ID	Date	PRCP	TMAX	TMIN
0	USW00003937	1970-05-02		10.3512	17.1943
1	USW00003937	1970-05-03		7.2561	27.5946
2	USW00003937	1970-05-04		13.547	26.2433
3	USW00003937	1970-05-05		12.5629	27.2496
4	USW00003937	1970-05-06		17.3708	32.3096

Figura 12 – Exemplo de visualização de dados originais e reconstruídos após execução do VAE.

4.1.4 Processamento Gráfico

Após a coleta, validação e (opcionalmente) reconstrução via VAE, os dados meteorológicos são analisados visualmente por meio de um conjunto de gráficos interativos. O sistema realiza automaticamente a geração dessas visualizações para **todas as variáveis numéricas consideradas válidas**, aplicando os mesmos procedimentos de análise para cada uma de forma individual e sobreposta, conforme aplicável.

As variáveis disponíveis são apresentadas dinamicamente por meio de menus interativos, nos quais o usuário pode selecionar, expandir e explorar os gráficos relacionados a cada uma delas. Esse mecanismo garante uma análise abrangente e personalizada, adaptando-se aos dados disponíveis e facilitando a interpretação visual em diferentes contextos.

4.1.4.1 Séries Temporais Interativas

Representam a evolução das variáveis ao longo do tempo. São úteis para identificação de tendências, sazonalidades e rupturas.

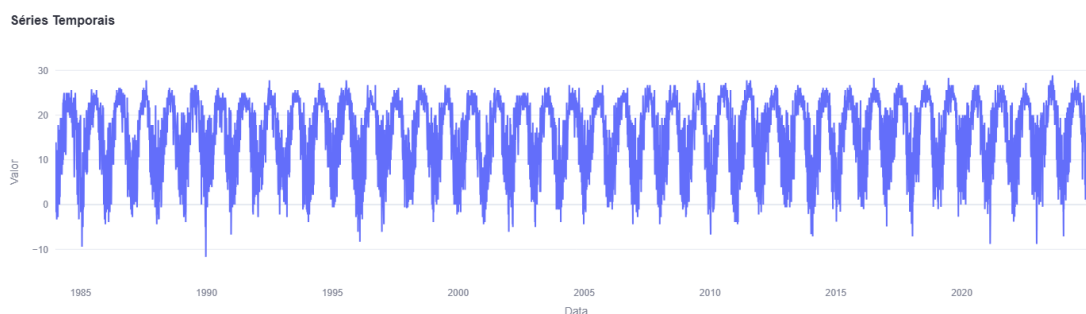


Figura 13 – Série temporal da variável TMIN.

4.1.4.2 Histograma com Curva de Densidade (KDE)

Permite observar a distribuição empírica de frequência de uma variável contínua, com sobreposição de uma estimativa suave da densidade de probabilidade.

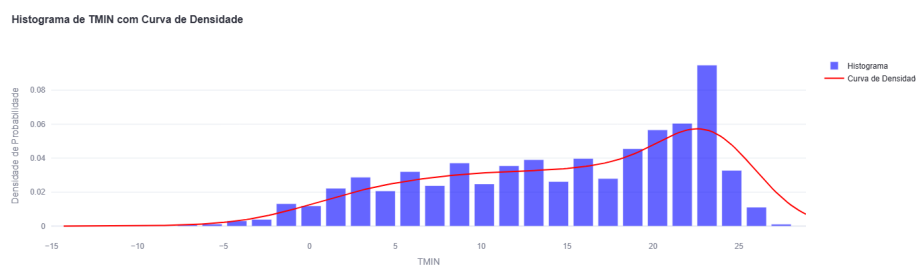


Figura 14 – Histograma com curva de densidade (KDE) para a variável TMIN.

4.1.4.3 Boxplots Interativos

Evidenciam mediana, quartis e outliers, permitindo comparação direta entre conjuntos e identificação de dispersões.

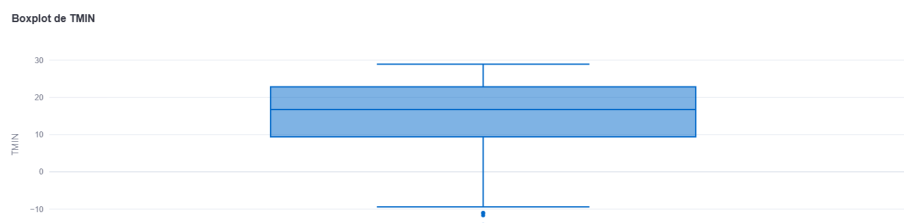


Figura 15 – Boxplot da variável TMIN com destaque para outliers.

4.1.4.4 Correlogramas ACF e PACF

Os correlogramas de autocorrelação (ACF) e autocorrelação parcial (PACF) permitem identificar dependências temporais entre observações.

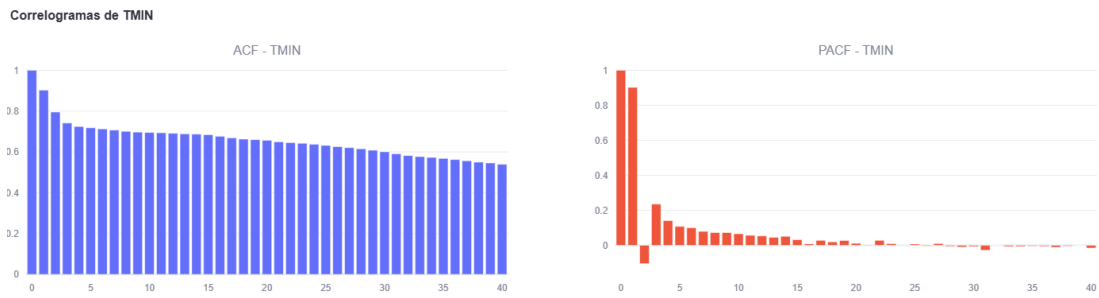


Figura 16 – Correlogramas ACF e PACF da variável TMIN.

4.1.4.5 Mapa de Calor Interativo de Correlação

Exibe uma matriz de correlação entre variáveis numéricas, facilitando a identificação de colinearidades e padrões conjuntos.

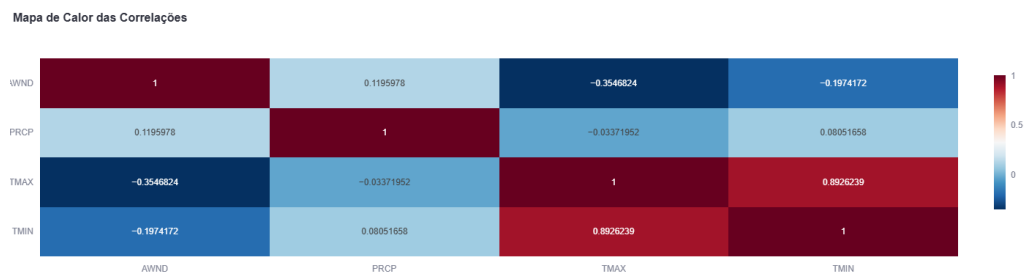


Figura 17 – Mapa de calor da matriz de correlação envolvendo a variável TMIN.

O sistema oferece um painel interativo que organiza as visualizações por variável numérica validada. Após o carregamento dos dados, são exibidas as estatísticas descritivas básicas (média, desvio padrão, quartis, mínimo e máximo), seguidas de menus recolhíveis que agrupam as visualizações específicas de cada variável. Essa organização facilita a navegação e a análise individualizada dos dados.

Coluna: PRCP - Válidos: 15066 / Nulos: 0
 Coluna: TMAX - Válidos: 15066 / Nulos: 0
 Coluna: TMIN - Válidos: 15066 / Nulos: 0

Estatísticas das Variáveis Válidas

	Variável	count	mean	std	min	25%	50%	75%	max
0	AWND	15066	34.2066	10.957	12.3194	25.9788	31.2465	40.9975	91.1366
1	PRCP	15066	37.4756	100.7065	-0.5072	11.6529	14.7579	21.2196	2583.8652
2	TMAX	15066	25.9039	5.9972	-2.1127	21.8477	27.1632	30.9705	35.3518
3	TMIN	15066	15.3996	6.5127	-5.4964	10.1725	17.0758	21.1696	24.7417

Visualizações

Visualizações para AWND

Visualizações para PRCP

Visualizações para TMAX

Visualizações para TMIN

Figura 18 – Painel interativo com estatísticas descritivas e visualizações individuais por variável numérica.

O conjunto de visualizações gerado pelo sistema permite avaliar, de forma visual e interativa, a aderência entre os dados originais e os reconstruídos via VAE. Para cada variável numérica validada, são produzidas dinamicamente representações gráficas — como séries temporais, histogramas, boxplots, correlogramas e mapas de calor — com sobreposição entre os dois conjuntos de dados. Essa abordagem proporciona uma comparação direta entre os padrões estatísticos e temporais dos dados brutos e dos dados reconstruídos, mesmo na ausência de uma métrica agregada consolidada.

A figura Figura 19 apresenta a sobreposição das séries temporais originais (*RawData*) e reconstruídas (VAE) para a variável TMIN, permitindo uma análise visual da similaridade entre os perfis das duas sequências ao longo do tempo.

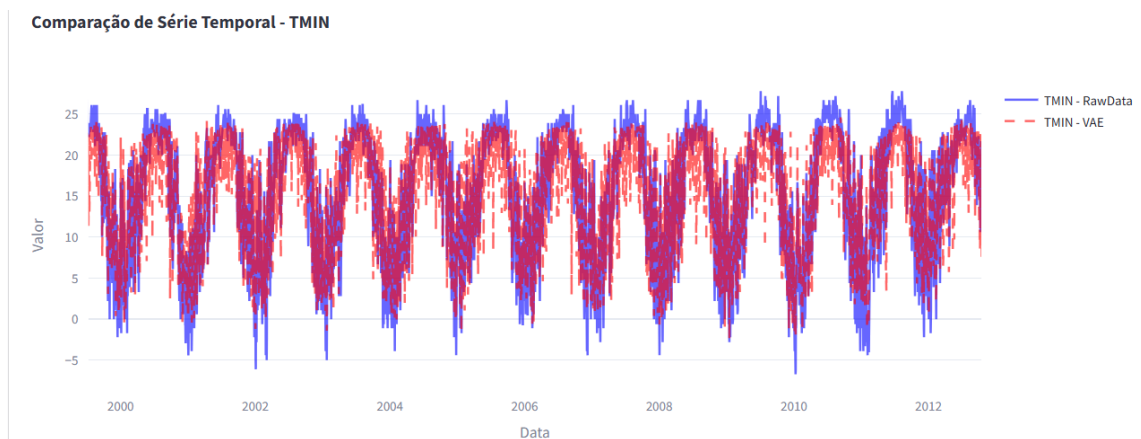


Figura 19 – Comparação de séries temporais para a variável TMIN: RawData (azul) e VAE (vermelho tracejado).

Adicionalmente, o sistema realiza o cálculo da distância DTW (*Dynamic Time Warping*) entre os pares de séries, fornecendo uma medida quantitativa da dissimilaridade entre os dados originais e os reconstruídos. Essa técnica é especialmente útil para séries temporais com possíveis desalinhamentos, pois permite o mapeamento não linear entre pontos semelhantes ao longo do tempo. A Figura 20 mostra o alinhamento resultante entre as duas versões da série TMIN, ilustrando como o VAE reproduz a estrutura sequencial dos dados originais.

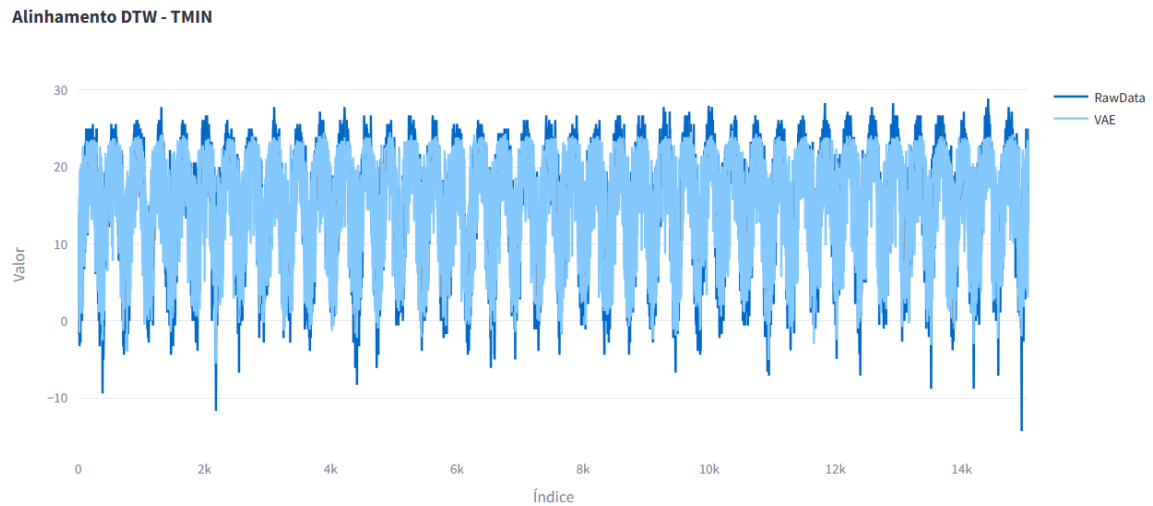


Figura 20 – Alinhamento entre séries RawData e VAE da variável TMIN utilizando a distância DTW.

5 DISCUSSÃO DOS RESULTADOS

Esta seção apresenta uma análise crítica dos resultados obtidos por meio do desenvolvimento do sistema proposto, destacando seus pontos fortes, limitações e aderência às boas práticas de Engenharia de Software. A reflexão fundamenta-se na aplicabilidade prática da solução no contexto da análise de séries temporais meteorológicas, bem como nos princípios técnicos adotados ao longo do processo de desenvolvimento.

Vantagens Identificadas

O sistema proposto apresenta diversas contribuições técnicas e operacionais relevantes:

- **Automatização de Etapas Analíticas:** A solução integra, de forma automatizada, as etapas de coleta, pré-processamento, validação, reconstrução e visualização de dados meteorológicos, eliminando a necessidade de execução manual de scripts isolados e reduzindo erros operacionais.
- **Interface Interativa baseada em Streamlit:** A escolha da biblioteca Streamlit permitiu o desenvolvimento ágil de uma interface gráfica amigável e responsiva, facilitando o uso por parte de pesquisadores com pouca familiaridade com programação. A usabilidade da ferramenta foi considerada satisfatória, permitindo a realização de análises complexas com mínima intervenção técnica.
- **Visualizações Avançadas com Plotly:** A implementação de gráficos interativos contribuiu para uma compreensão mais aprofundada das variáveis envolvidas, permitindo a análise exploratória de tendências, outliers, padrões sazonais e relações estatísticas entre variáveis.
- **Reprodutibilidade e Organização de Saídas:** Os resultados são sistematicamente armazenados em diretórios hierárquicos organizados por estação meteorológica, período analisado e tipo de dado (original ou reconstruído), facilitando reprocessamentos, auditorias e comparações subsequentes.
- **Reconstrução de Dados com Variational Autoencoder (VAE):** A utilização de um modelo generativo baseado em VAE representa um diferencial na reconstrução de valores ausentes, contribuindo para a consistência dos dados analisados. A abordagem permite não apenas a imputação de falhas, mas também a análise da estrutura latente das séries temporais.
- **Análises Estatísticas Completas:** O sistema contempla um conjunto robusto de métricas estatísticas descritivas (tendência central, dispersão, assimetria, curtose, autocorrelação, entre outras), promovendo uma caracterização detalhada das séries temporais processadas.

Limitações e Pontos de Melhoria

Apesar dos avanços apresentados, algumas limitações foram identificadas:

- **Dependência de Conectividade:** A coleta de dados diretamente dos servidores do NOAA pressupõe disponibilidade constante de conexão com a internet e integridade do serviço remoto, o que pode comprometer a continuidade do processo em ambientes restritos.
- **Custo Computacional do VAE:** A reconstrução de dados via VAE exige recursos computacionais significativos, especialmente em contextos com grandes volumes de dados ou ausência de aceleração por GPU. Isso pode impactar o tempo de resposta em ambientes com infraestrutura limitada.
- **Ajuste Manual de Hiperparâmetros:** A calibração do VAE é sensível à configuração de hiperparâmetros como taxa de aprendizado, tamanho do lote e dimensão latente. A necessidade de ajustes manuais pode restringir o uso pleno da ferramenta por usuários com menor domínio de aprendizado de máquina.
- **Foco Univariado em Algumas Visualizações:** Algumas representações gráficas, como histogramas e boxplots, trabalham com variáveis isoladas. Isso limita a exploração de correlações multivariadas e a aplicação de modelos preditivos baseados em interdependência entre variáveis.
- **Escalabilidade Limitada:** O sistema foi inicialmente projetado para execução local e uso individual. Para ambientes de múltiplos usuários ou implantação em servidores web, seriam necessárias adaptações arquiteturais, como autenticação, paralelização e suporte à concorrência.

Aderência à Engenharia de Software

O projeto foi conduzido à luz de princípios fundamentais da Engenharia de Software, com o intuito de garantir qualidade estrutural, manutenibilidade e extensibilidade da aplicação. Os principais aspectos observados incluem:

- **Modularidade:** O sistema foi estruturado em módulos funcionais independentes (por exemplo, `extracao.py`, `vae.py`, `graphing.py`), com clara separação de responsabilidades. Essa organização segue o princípio da responsabilidade única (SRP), promovendo baixo acoplamento e alta coesão.
- **Reusabilidade de Componentes:** Funções como `extrair_dados_validos` e `calculate_sta` foram concebidas para uso genérico, facilitando a reutilização em diferentes contextos analíticos e favorecendo a evolução do sistema.

- **Separação entre Lógica e Interface:** A lógica de processamento foi isolada da camada de apresentação, permitindo mudanças na interface do usuário sem comprometer a lógica computacional, e vice-versa.
- **Facilidade de Uso e Acessibilidade:** A aplicação da filosofia de design centrado no usuário (User-Centered Design) viabilizou o uso por perfis variados de usuários, democratizando o acesso às ferramentas analíticas propostas.
- **Organização e Rastreabilidade de Arquivos:** O armazenamento padronizado dos resultados, por estação e período, favorece a rastreabilidade dos dados e o controle de versões de experimentos.
- **Escalabilidade Potencial:** A arquitetura adotada permite extensões futuras, como implantação em contêineres Docker, integração com bancos de dados meteorológicos externos via APIs e disponibilização da solução em nuvem.
- **Manutenibilidade e Testabilidade:** A separação modular permite substituições pontuais de componentes (como o modelo de reconstrução) sem impacto sistêmico. A estrutura adotada também favorece o desenvolvimento de testes unitários e de integração.

Síntese Final

Pode-se concluir que o sistema desenvolvido atende de forma satisfatória aos objetivos propostos, conciliando técnicas modernas de ciência de dados com práticas consolidadas da Engenharia de Software. A solução representa não apenas um avanço em termos de automação e qualidade analítica, mas também um exemplo prático da aplicação de princípios de arquitetura de sistemas em um projeto acadêmico, contribuindo para a formação técnica e científica do discente envolvido.

6 CONCLUSÃO

O presente trabalho teve como objetivo o desenvolvimento de um sistema interativo para coleta, tratamento, visualização e comparação de dados meteorológicos, com ênfase na análise descritiva e na reconstrução de séries temporais por meio de técnicas de aprendizado profundo, em especial o *Variational Autoencoder* (VAE). A ferramenta construída permite que o usuário percorra todas as etapas da análise — desde a seleção da estação e do intervalo temporal até a validação, visualização gráfica e comparação estatística — de forma simples e intuitiva, promovendo uma experiência acessível mesmo para indivíduos sem conhecimentos avançados em programação ou ciência de dados. As funcionalidades oferecidas abrangem desde estatísticas descritivas básicas até testes de estacionariedade e análises de correlação temporal, organizadas em uma interface amigável, desenvolvida com as bibliotecas *Streamlit* e *Plotly*.

A principal contribuição deste projeto reside na integração de múltiplas etapas do fluxo de análise climatológica em um ambiente único, interativo e personalizável. Diferentemente de soluções fragmentadas ou ferramentas genéricas para análise de séries temporais, o sistema desenvolvido permite não apenas visualizar os dados brutos, mas também validá-los automaticamente e compará-los com versões reconstruídas por modelos generativos. Essa abordagem amplia o potencial de investigação científica e a tomada de decisões orientadas por dados, especialmente em contextos em que a ausência de informações ou a presença de ruídos nas observações representa um desafio analítico relevante.

Como direções futuras, propõe-se a ampliação do sistema com funcionalidades voltadas à modelagem preditiva, incorporação de métodos de imputação alternativos ao VAE (como redes LSTM ou interpoladores híbridos), e suporte a análises multivariadas mais robustas. Além disso, a adaptação do sistema para ambientes distribuídos — como aplicações web multiusuário ou serviços em nuvem — poderá ampliar significativamente seu alcance e aplicabilidade em projetos de monitoramento ambiental, agricultura de precisão e estudos de mudanças climáticas em larga escala.

REFERÊNCIAS

- BERNDT, D. J.; CLIFFORD, J. Using dynamic time warping to find patterns in time series. In: **Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining**. Seattle: AAAI Press, 1994. p. 359–370.
- HUNTER, J. D. Matplotlib: A 2d graphics environment. **Computing in Science & Engineering**, IEEE, v. 9, n. 3, p. 90–95, 2007.
- Intergovernmental Panel on Climate Change. **Climate Change 2021: The Physical Science Basis**: Summary for policymakers. Geneva, 2021. Disponível em: <https://www.ipcc.ch/report/ar6/wg1/>. Acesso em: 15 jan. 2025.
- KAUR, S.; BANSAL, M. Real-time fault detection in climate sensors using deep autoencoders. **Environmental Data Science**, Cambridge University Press, Cambridge, v. 1, p. e11, 2023.
- KINGMA, D. P.; WELLING, M. **Auto-Encoding Variational Bayes**. 2014. E-print arXiv:1312.6114. Disponível em: <https://arxiv.org/abs/1312.6114>. Acesso em: 10 ago. 2024.
- Max Planck Institute for Meteorology. **CDO User Guide: Climate Data Operators**. Hamburg, 2020. Disponível em: <https://code.mpimet.mpg.de/projects/cdo>. Acesso em: 10 fev. 2024.
- MCKINNEY, W. Data structures for statistical computing in python. In: WALT, S. van der; MILLMAN, J. (Ed.). **Proceedings of the 9th Python in Science Conference**. Austin: [s.n.], 2010. p. 56–61.
- PEREIRA, A.; OLIVEIRA, J. Aplicação de autoencoders variacionais na previsão de chuvas. In: **Anais do XVI Congresso Brasileiro de Meteorologia**. Maceió: SBMET, 2020. Disponível em: <https://eventos.mcti.gov.br/cbm2020>. Acesso em: 12 mar. 2024.
- Plotly Technologies Inc. **Plotly: Collaborative Data Science**. Montreal: Plotly Technologies Inc., 2015. Software library. Disponível em: <https://plotly.com>. Acesso em: 10 ago. 2024.
- SEABOLD, S.; PERKTOLD, J. Statsmodels: Econometric and statistical modeling with python. In: **Proceedings of the 9th Python in Science Conference**. Austin: [s.n.], 2010. p. 92–96.
- STORCH, H. von; ZWIERS, F. W. **Time Series Analysis and Applications in Climate Science**. Cambridge: Cambridge University Press, 2006.
- Streamlit Inc. **Streamlit: The fastest way to build data apps in Python**. San Francisco, 2023. Disponível em: <https://streamlit.io>. Acesso em: 01 out. 2024.
- World Meteorological Organization. **CLIMSOFT User Guide**. Geneva, 2022. Disponível em: <https://public.wmo.int/en/resources/library/climsoft>. Acesso em: 20 set. 2024.
- ZHANG, X.; YANG, F. **RClimDex (1.0): Software for Climate Extremes Indices**. Downsview, 2009. User Manual. Disponível em: <https://etccdi.pacificclimate.org/software.shtml>. Acesso em: 05 nov. 2023.
- ZHOU, L.; CHEN, Y.; WANG, J. A hybrid cnn-lstm model for multivariate climate time series forecasting. **Journal of Climate Informatics**, v. 8, n. 2, p. 45–58, 2022.