



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA**

JEFFERSON VIEIRA DOS SANTOS

**UM ESTUDO DO *STANDARDIZED PRECIPITATION INDEX* (SPI) E TÉCNICAS DE
MACHINE LEARNING PARA PREVISÃO DE SECAS NO ESTADO DA PARAÍBA**

CAMPINA GRANDE - PB

2023

JEFFERSON VIEIRA DOS SANTOS

UM ESTUDO DO *STANDARDIZED PRECIPITATION INDEX* (SPI) E TÉCNICAS DE MACHINE LEARNING PARA PREVISÃO DE SECAS NO ESTADO DA PARAÍBA

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Área de concentração: Estatística .

Orientador: Prof. Dr. Silvio Fernando Alves Xavier Júnior.

**CAMPINA GRANDE - PB
2023**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S237e Santos, Jefferson Vieira dos.

Um estudo do Standardized Precipitation Index (SPI) e técnicas de machine learning para previsão de secas no estado da Paraíba [manuscrito] / Jefferson Vieira dos Santos. - 2023.
45 p.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2023.

"Orientação : Prof. Dr. Prof. Dr. Silvio Fernando Alves Xavier Júnior, Departamento de Estatística - CCT. "

1. Índice Padronizado de Precipitação (SPI). 2. Seca. 3. Machine Learning. I. Título

21. ed. CDD 519.5

JEFFERSON VIEIRA DOS SANTOS

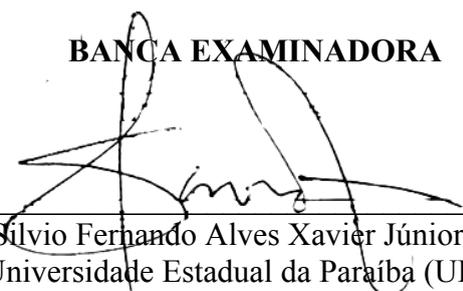
UM ESTUDO DO *STANDARDIZED PRECIPITATION INDEX* (SPI) E TÉCNICAS DE MACHINE LEARNING PARA PREVISÃO DE SECAS NO ESTADO DA PARAÍBA

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Área de concentração: Estatística .

Aprovado em: 14 / 11 / 2023 .

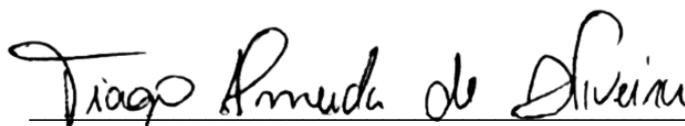
BANCA EXAMINADORA



Prof. Dr. Sílvio Fernando Alves Xavier Júnior. (Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Elias Dias Coelho Neto
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba (UEPB)

RESUMO

Objetivou-se neste estudo identificar e analisar as secas na Paraíba, utilizando o Índice Padronizado de Precipitação (SPI) e técnicas de modelagem com algoritmos de machine learning para prever o SPI para os anos subsequentes (2020-2021) em seis estações pluviométricas distribuídas nas mesorregiões da Paraíba. Os dados de precipitação foram obtidos a partir do Global Precipitation Climatology Centre (GPCC) e da base de dados da National Oceanic and Atmospheric Administration (NOAA), abrangendo o período de 1991 a 2019. Foram selecionados três algoritmos de machine learning com base em sua capacidade de ajuste aos dados históricos de SPI: Extra Trees Regressor, Gradient Boosting Regressor e Random Forest Regressor. Os modelos de machine learning aplicados apresentaram resultados satisfatórios, com destaque para o Extra Trees Regressor, que consistentemente produziu o maior valor de R^2 em todas as estações, indicando uma alta explicabilidade dos dados. As previsões foram analisadas para determinar sua precisão e confiabilidade, fornecendo insights valiosos sobre a variabilidade da precipitação e a ocorrência de secas nas diferentes mesorregiões da Paraíba. Em conclusão, este estudo contribuiu para a compreensão da variabilidade climática e de suas implicações na Paraíba, fornecendo insights valiosos sobre a ocorrência de secas e a importância de abordagens adaptativas para mitigar impactos adversos. A aplicação do SPI e técnicas de machine learning mostrou-se eficaz na análise e previsão da precipitação, oferecendo uma abordagem objetiva para caracterizar a intensidade das secas e chuvas em determinadas regiões.

Palavras-Chave: índice padronizado de precipitação; seca; machine learning.

ABSTRACT

The aim of this study was to identify and analyze droughts in Paraíba, using the Standardized Precipitation Index (SPI) and machine learning algorithms for predicting SPI for the subsequent years (2020-2021) at six rainfall stations distributed across the mesoregions of Paraíba. Precipitation data were obtained from the Global Precipitation Climatology Centre (GPCC) and the National Oceanic and Atmospheric Administration (NOAA) database, covering the period from 1991 to 2019. Three machine learning algorithms were selected based on their ability to fit historical SPI data: Extra Trees Regressor, Gradient Boosting Regressor, and Random Forest Regressor. The applied machine learning models yielded satisfactory results, with Extra Trees Regressor consistently producing the highest R^2 value across all stations, indicating high data explainability. The predictions were analyzed to determine their accuracy and reliability, providing valuable insights into precipitation variability and drought occurrence in different mesoregions of Paraíba. In conclusion, this study contributed to understanding climate variability and its implications in Paraíba, offering valuable insights into drought occurrence and the importance of adaptive approaches to mitigate adverse impacts. The application of SPI and machine learning techniques proved to be effective in analyzing and predicting precipitation, providing an objective approach to characterizing drought and rainfall intensity in specific regions.

Keywords: standardized precipitation index; drought; machine learning.

LISTA DE ILUSTRAÇÕES

Figura 1 – Histogramas das precipitações de chuvas.....’	19
Figura 2 – Séries temporais das precipitações de chuvas entre 1991-01-01 à 2019-12-01.....	20
Figura 3 – SPI-1.....	22
Figura 4 – SPI-3.....	23
Figura 5 – SPI-6.....	24
Figura 6 – SPI-9.....	26
Figura 7 – SPI-12.....	27
Figura 8 – SPI-24.....	29
Figura 9 – Gráficos de linha das 6 estações referente às previsões do SPI 9.....	37
Figura 10 – Gráficos de linha das 6 estações referente às previsões do SPI 12.....	39
Figura 11 – Gráficos de linha das 6 estações referente às previsões do SPI 24.....	41

LISTA DE TABELAS

Tabela 1 –	Classificação do índice de Precipitação padronizada.....	13
Tabela 2 –	Estatística Descritiva das Precipitações de chuvas entre 1991-01-01 à 2019-12-01.....	17
Tabela 3 –	SPI-1.....	21
Tabela 4 –	SPI-3.....	22
Tabela 5 –	SPI-6.....	24
Tabela 6 –	SPI-9.....	25
Tabela 7 –	SPI-12.....	26
Tabela 8 –	SPI-24.....	28
Tabela 9 –	Comparativo de Desempenho dos Modelos de Aprendizado de Máquina para o SPI-9.....	29
Tabela 10 –	Comparativo de Desempenho dos Modelos de Aprendizado de Máquina para o SPI-12.....	31
Tabela 11 –	Comparativo de Desempenho dos Modelos de Aprendizado de Máquina para o SPI-12.....	32

LISTA DE ABREVIATURAS E SIGLAS

ET	Extra Trees Regressor
GB	Gradient Boosting Regressor
GPCC	Global Precipitation Climatology Centre
MAE	Erro Médio Absoluto
MAPE	Erro Percentual Absoluto Médio
MSE	Erro Médio Quadrático
NOAA	National Oceanic and Atmospheric Administration
OMM	Organização Mundial de Meteorologia
RF	Random Forest Regressor
RMSE	Raiz do Erro Médio Quadrático
RMSLE	Raiz do Erro Logarítmico Médio Quadrático
SPI	Standardized Precipitation Index

SUMÁRIO

1	INTRODUÇÃO	9
2	METODOLOGIA	11
2.1	Dados utilizados	11
2.2	Regiões e seus climas	11
2.3	Índice de Precipitação Padronizado (SPI)	11
2.4	Modelagem	14
2.4.1	<i>Extra Trees Regressor (Árvores Extremamente Aleatórias)</i>	15
2.4.2	<i>Random Forest Regressor (Árvores Aleatórias)</i>	15
2.4.3	<i>Gradient Boosting Regressor</i>	16
2.4.4	<i>Crítérios para avaliação do modelo</i>	17
3	RESULTADOS E DISCUSSÕES	18
3.1	Análise descritiva	18
3.2	Standardized Precipitation Index (SPI)	20
3.3	Modelagem	29
3.4	Previsões	35
4	CONCLUSÃO	42
	REFERÊNCIAS	43

1 INTRODUÇÃO

A compreensão e previsão de eventos hidrometeorológicos são de extrema importância tanto para a comunidade científica quanto para a sociedade em geral. Eventos extremos de precipitação, como secas e enchentes, possuem impactos significativos em diversos setores econômicos e sociais, tornando essencial o desenvolvimento de métodos precisos e confiáveis para sua avaliação. No entanto, o uso de técnicas estatísticas convencionais muitas vezes não é suficiente para modelar e prever esses fenômenos complexos, devido aos erros de estimação inerentes a esses métodos (HASAN et al., 2012).

Dentre os fenômenos hidrometeorológicos, a seca assume um papel de destaque devido à sua recorrência e complexidade. Além de influenciar diretamente as atividades agrícolas, a seca tem implicações socioeconômicas, ambientais e políticas, afetando diversas regiões do mundo. Em especial, regiões áridas e semiáridas são particularmente vulneráveis aos efeitos devastadores da seca (Mathbout et al., 2018; Wang et al., 2014). No contexto brasileiro, a região Nordeste, notadamente a Paraíba, enfrenta períodos prolongados de seca e estiagem, motivando a realização de pesquisas climáticas e meteorológicas na região.

A intersecção entre as áreas de estatística e meteorologia tem se mostrado fundamental para a análise desses fenômenos complexos. A aplicação de métodos estatísticos, como o Índice Padronizado de Precipitação, tem se destacado como uma ferramenta crucial na análise da variabilidade climática, principalmente no contexto da seca. O SPI, ao quantificar os desvios da precipitação em diferentes escalas temporais, oferece uma abordagem objetiva para caracterizar a intensidade das secas e chuvas em determinadas regiões (Macedo et al., 2010; Macedo et al., 2011; De Sousa Guedes et al., 2012; De Farias et al., 2014; Santos et al., 2019; Da Silva et al., 2021).

A interdisciplinaridade entre estatística e meteorologia, combinada com o domínio de ferramentas computacionais como R e Python, se mostra essencial para a formação de pesquisadores capazes de enfrentar os desafios climáticos atuais.

Neste contexto, este trabalho teve como objetivo aplicar o Índice Padronizado de Precipitação (SPI) e fazer modelagem com técnicas e algoritmos de machine learning para previsão do SPI para o ano subsequente (2020-2021) nas seis estações pluviométricas distribuídas nas mesorregiões da Paraíba. Cada estação representa uma mesorregião específica do estado, a saber: Mata Paraibana (João Pessoa), Agreste Paraibano (Campina Grande), Borborema (Monteiro e Picuí) e Sertão Paraibano (Patos e Sousa). A análise do SPI

nessas diferentes mesorregiões permitirá uma compreensão mais detalhada da variabilidade da precipitação e da ocorrência de secas em escalas temporais e geográficas diversas.

2 METODOLOGIA

2.1 Dados utilizados

Os dados utilizados neste estudo foram obtidos a partir do Global Precipitation Climatology Centre (GPCC) (<https://psl.noaa.gov/data/gridded/data.gpcc.html>), um renomado repositório amplamente utilizado em pesquisas climáticas, e da base de dados extensa da National Oceanic and Atmospheric Administration (NOAA). Essas fontes de dados são conhecidas por sua confiabilidade e abrangência, proporcionando informações valiosas sobre as condições climáticas e as variações de precipitação ao longo do tempo. As séries temporais de dados de precipitação abrangeram o período de 1991 a 2019, permitindo uma análise abrangente das tendências climáticas ao longo dos anos.

2.2 Regiões e seus climas

As mesorregiões selecionadas para este estudo na Paraíba exibem uma rica diversidade climática, influenciada por sua localização geográfica e características topográficas. A Mata Paraibana, representada pela cidade de João Pessoa, é caracterizada por um clima tropical úmido com chuvas bem distribuídas ao longo do ano. O Agreste Paraibano, representado por Campina Grande, possui um clima tropical com estação seca bem definida e uma estação chuvosa. A região da Borborema, representada por Monteiro e Picuí, apresenta um clima tropical semiárido com chuvas escassas e irregulares. Por fim, o Sertão Paraibano, representado por Patos e Sousa, é caracterizado por um clima semiárido, com chuvas ainda mais escassas e concentradas em um curto período (Paraíba Total, 2021). Essas diferentes condições climáticas entre as mesorregiões resultam em variabilidades notáveis nos padrões de precipitação ao longo do ano.

2.3 Índice de Precipitação Padronizado (SPI)

O Índice Padronizado de Precipitação (SPI) é uma ferramenta estatística amplamente empregada para a avaliação de condições de seca ou excesso de chuvas em uma região específica. Esse índice se fundamenta na análise da distribuição histórica da precipitação ao longo do tempo e no desvio padrão dos dados pluviométricos passados. Por meio do SPI, é possível quantificar a intensidade, duração e frequência de eventos de escassez hídrica ou chuvas intensas. Originalmente desenvolvido por MCKEE et al. (1993), o SPI visa a medir o déficit ou excesso de precipitação em distintas escalas temporais (1, 3, 6, 12, 24 e 48 meses).

A adoção dessa metodologia é amplamente respaldada pela Organização Meteorológica Mundial (OMM).

O SPI é expresso em desvios padrão que representam as discrepâncias entre os registros de precipitação observados e as médias de longo prazo, quando assumida uma distribuição normal. Dado que a precipitação não segue uma distribuição normal, uma transformação é aplicada para, em seguida, ajustar os dados a uma distribuição normalizada de probabilidade. Esse processo de normalização torna possível comparar climas mais secos e úmidos de maneira equitativa, habilitando, assim, a análise tanto de períodos chuvosos quanto secos (FERNANDES et al., 2009).

A aplicação do SPI requer a disponibilidade de dados pluviométricos ao longo de pelo menos 30 anos. O cálculo desse índice baseia-se nas características estatísticas das séries históricas de precipitação em diferentes intervalos temporais, empregando conceitos de densidade de probabilidade. Para ajustar a distribuição de frequência do total de precipitação em uma série pluviométrica específica, foram consideradas distribuições estatísticas como a Normal e a Gama.

A distribuição de probabilidade Gama é representada pela seguinte função:

$$g(x) = \frac{1}{\beta^\alpha \delta(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad (1)$$

Na qual $\alpha > 0$ representa o parâmetro de forma, $b > 0$ é o parâmetro de escala e $x > 0$ é a quantidade de precipitação pluviométrica e $\delta(\alpha)$ é a função Gama dada pela seguinte expressão:

$$\delta(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \quad (2)$$

Na fórmula mencionada, a variável temporária y pode variar no intervalo de 0 a ∞ . Os coeficientes são determinados por meio do método de máxima verossimilhança, conforme delineado por THOM (1966), e são calculados utilizando as seguintes equações:

$$A = \ln \ln(\bar{x}) - \frac{\sum_i^n x_i}{n} \quad (3)$$

$$\hat{\alpha} = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \quad (4)$$

$$\hat{b} = \frac{\bar{x}}{\hat{\alpha}} \quad (5)$$

O tamanho da amostra é representado por n , e o valor médio dos dados de precipitação é denotado como x . A função de distribuição acumulada da distribuição Gama é definida da seguinte maneira:

$$G(x) = \frac{1}{\delta(\hat{\alpha})} \int_0^x t^{\hat{\alpha}-1} e^{-t} dt \quad (6)$$

A função de densidade de probabilidade gama, g , não é definida para $x=0$. No entanto, considerando que a amostra de dados de precipitação pode incluir valores iguais a zero, a probabilidade cumulativa é expressa da seguinte maneira:

$$H(x) = q + (1 - q)G(x) \quad (7)$$

Nesta equação, "q" representa a probabilidade de ocorrência de um valor zero, a qual é calculada como $q = m/n$. Nesse contexto, "m" representa o número de eventos em que a precipitação foi igual a zero, enquanto "n" é o número total de observações. Ao transformar a probabilidade cumulativa $H(x)$ em uma variável aleatória padronizada (Z), com média zero e desvio padrão de um, podemos simplificar a análise estatística. A variável "Z" corresponde ao valor do Índice Padronizado de Precipitação (SPI). Segundo Abramowitz e Stegun (1965), a relação entre as distribuições de probabilidade Gama e Normal é representada por:

$$Z = - \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad 0 < H(x) \leq 0,5 \quad (8)$$

$$Z = + \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad 0,5 < H(x) \leq 1 \quad (9)$$

$$t = \sqrt{\ln \ln \left(\frac{1}{(H(x))^2} \right)} \quad 0 < H(x) \leq 0,5, \quad (10)$$

$$t = \sqrt{\ln \ln \left(\frac{1}{(1-H(x))^2} \right)} \quad 0,5 < H(x) \leq 1 \quad (11)$$

Para os valores de coeficiente iguais a: $c_0 = 2,515517$; $c_1 = 0,802853$; $c_2 = 0,010328$;
 $d_1 = 1,432788$; $d_2 = 0,189269$; $d_3 = 0,001308$

Para iniciar o cálculo do Índice Padronizado de Precipitação (SPI), é necessário primeiramente ajustar a função densidade de probabilidade Gama às séries de totais mensais de precipitação. Em seguida, a probabilidade cumulativa de ocorrência de cada total mensal é estimada. Posteriormente, aplica-se a função Normal e a inversa da distribuição Gaussiana a essa probabilidade, a fim de se obter o valor do SPI correspondente.

O evento de seca é identificado quando o valor do SPI se torna negativo e atinge a marca de -1. Esse evento se encerra quando o SPI retorna a valores positivos, de acordo com a metodologia proposta por MCKEE et al. (1993). Na escala do índice, valores menores ou iguais a -2 são indicativos de "seca extrema", enquanto valores maiores ou iguais a +2 sinalizam condições de "umidade extrema".

Tabela 1- Classificação do índice de Precipitação Padronizada.

Valores de SPI	Classificação
$SPI \geq 2,00$	Extremamente úmido
$1,00 \geq SPI > 2,00$	Muito úmido
$0,50 \geq SPI > 1,00$	Moderadamente úmido
$-0,50 > SPI > 0,50$	Próximo ao normal
$-1,00 > SPI \geq -0,50$	Moderadamente seco
$-2,00 > SPI \geq -1,00$	Muito seco
$SPI \leq -2,00$	Extremamente seco

Fonte: FERNANDES et al. (2009)

2.4 Modelagem

O estudo foi estruturado em várias etapas para garantir uma análise abrangente e precisa da previsão do Índice Padronizado de Precipitação (SPI) para os anos de 2020 e 2021 nas seis estações pluviométricas distribuídas nas mesorregiões da Paraíba. Inicialmente, foram coletados os dados históricos de SPI nas seis estações pluviométricas, considerando as escalas temporais de 9, 12 e 24 meses. Esses dados serviram como base para o treinamento e teste dos modelos de machine learning. Utilizando a biblioteca PyCaret, três algoritmos de machine learning foram selecionados com base em sua capacidade de ajuste aos dados históricos de SPI. Os algoritmos escolhidos foram: Extra Trees Regressor, Gradient Boosting Regressor e Random Forest Regressor. Após o treinamento e ajuste, os modelos foram aplicados para prever o SPI para o ano de 2020. As previsões foram então analisadas para determinar sua precisão e confiabilidade, fornecendo insights valiosos sobre a variabilidade

da precipitação e a ocorrência de secas nas diferentes mesorregiões da Paraíba. A seguir, são apresentados de maneira mais detalhada os modelos empregados neste estudo.

2.4.1 Extra Trees Regressor (*Árvores Extremamente Aleatórias*)

O algoritmo Extra Trees, ou Árvores Extremamente Aleatorizadas, é um método de aprendizado de máquina que pertence à categoria de algoritmos de florestas aleatórias. Este algoritmo é particularmente útil para tarefas de regressão e classificação, oferecendo uma abordagem mais aleatória na criação de árvores de decisão (Geurts, 2006).

Em cada nó da árvore, uma seleção aleatória de um determinado número de atributos é feita e, entre eles, o melhor é escolhido. No extremo, o método opta por um único atributo e ponto de corte de forma aleatória em cada nó, resultando na construção de árvores totalmente randomizadas cujas estruturas não dependem dos valores da variável-alvo do conjunto de treinamento, conforme ressaltado por Geurts (2006). Ou seja, no algoritmo Extra Trees, a seleção e divisão dos nós é efetuada de forma ainda mais aleatória em relação a outros métodos de florestas aleatórias, daí a origem do termo "extremamente" em seu nome. Após a seleção aleatória das variáveis candidatas para o nó inicial, os valores existentes para cada variável são divididos ao acaso. Esse procedimento é replicado em cada subsequente nó filho até que um nó folha seja alcançado. As previsões de todas as árvores são então combinadas para definir a previsão final, baseada na votação da maioria.

De acordo com Geurts (2006), uma das principais vantagens do Extra Trees é a sua capacidade de melhorar a precisão do modelo devido à sua aleatoriedade adicional. Além disso, ele é eficaz na redução do overfitting, proporcionando modelos mais generalizados e robustos para diversas tarefas de aprendizado de máquina.

2.4.2 Random Forest Regressor (*Árvores Aleatórias*)

O algoritmo Random Forest é uma classe de métodos de conjunto projetados especificamente para classificadores de árvores de decisão. Ela combina as previsões feitas por múltiplas árvores de decisão, onde cada árvore é gerada com base nos valores de um conjunto independente de vetores aleatórios de acordo com Tan, Steinbach e Kumar (2005).

Tan, Steinbach e Kumar (2005) descrevem que as florestas aleatórias utilizam vetores aleatórios gerados de uma distribuição de probabilidade fixa para construir árvores de decisão, onde a seleção de características e a combinação de previsões são feitas de maneira a reduzir o viés e aumentar a robustez do modelo. Eles também discutem diferentes estratégias

para incorporar aleatoriedade no processo de crescimento das árvores e como a escolha do número de características influencia a correlação e a força das árvores dentro da floresta. Além disso, os autores apontam que a eficácia das florestas aleatórias em termos de precisão de classificação é comparável ao AdaBoost, com vantagens adicionais de robustez e velocidade.

Em termos práticos, nas Random Forest, durante o treinamento, cada árvore é desenvolvida por meio da seleção aleatória de amostras do conjunto de dados, usando reposição. Em cada nó da árvore, um subconjunto aleatório de características é escolhido para decidir a divisão. Esta abordagem aleatória na formação das árvores amplia a diversidade entre elas, potencializando a robustez e acurácia do modelo resultante. Posterior ao treinamento, para novas entradas, cada árvore na floresta realiza uma previsão. A decisão final é feita pela maioria dos votos das árvores em tarefas de classificação ou pela média das previsões em tarefas de regressão.

2.4.3 Gradient Boosting Regressor

O Gradient Boosting Regressor é um algoritmo de aprendizado de máquina poderoso e eficiente que utiliza a técnica de boosting para otimizar a precisão das previsões em tarefas de regressão. Ele constrói múltiplos modelos de previsão de maneira sequencial, onde cada modelo tenta corrigir os erros do seu antecessor.

Semelhante ao Random Forest, o Gradient Boosting Regressor (GBR) (Friedman, 2001) é uma técnica de aprendizado de máquina também baseada na média das árvores. No entanto, em vez de treinar muitas árvores de alta variância completas que são médias para evitar o sobreajuste, o GBR adiciona sequencialmente pequenas árvores, cada uma com alto viés. Em cada iteração, a nova árvore a ser adicionada foca explicitamente nos documentos que são responsáveis pelo erro de regressão remanescente atual.

O funcionamento do Gradient Boosting Regressor inicia com um modelo base, que pode ser um modelo simples, e calcula os resíduos (diferenças entre as previsões e os valores reais). Em seguida, um novo modelo é construído para prever esses resíduos. Este processo é repetido várias vezes, e cada novo modelo é adicionado ao modelo global para melhorar a precisão das previsões. O algoritmo ajusta-se iterativamente para minimizar o erro residual, otimizando a capacidade de previsão do modelo final.

2.4.4 Critérios para avaliação do modelo

O Erro Médio Absoluto (MAE), Erro Médio Quadrático (MSE), Raiz do Erro Médio Quadrático (RMSE), Coeficiente de Determinação (R^2), Raiz do Erro Logarítmico Médio Quadrático (RMSLE) e o Erro Percentual Absoluto Médio (MAPE) foram utilizados para a avaliação dos resultados. As seguintes equações resumem essas medidas.

$$MAE = \left(\frac{1}{n}\right) \Sigma |y_i - \hat{y}_i| \quad (12)$$

$$MSE = \left(\frac{1}{n}\right) \Sigma (y_i - \hat{y}_i)^2 \quad (13)$$

$$RMSE = \text{sqr}t\left[\left(\frac{1}{n}\right) \Sigma (y_i - \hat{y}_i)^2\right] \quad (14)$$

$$R^2 = 1 - \left(\frac{SSE}{SST}\right) \quad (15)$$

$$RMSLE = \text{sqr}t\left[\left(\frac{1}{n}\right) \Sigma (\log \log (y_i + 1) - \log \log (\hat{y}_i + 1))^2\right] \quad (16)$$

$$MAPE = \left(\frac{1}{n}\right) \Sigma \left| \frac{(y_i - \hat{y}_i)}{y_i} \right| * 100 \quad (17)$$

onde y_i , \bar{y} e \hat{y}_i são, respectivamente, a i -ésima observação, o valor médio dos dados de observação e o valor previsto para o i -ésimo dado e SSE é a soma dos quadrados dos resíduos e SST é a soma total dos quadrados.

3 RESULTADOS E DISCUSSÕES

3.1 Análise descritiva

A tabela 2 apresenta informações sobre as precipitações de chuva em seis estações pluviométricas, incluindo dados geográficos e estatísticos. João Pessoa é a cidade com a maior média de chuvas, registrando em média 130 milímetros, enquanto Picuí tem a menor média, com 43 milímetros. É importante destacar que João Pessoa e Sousa apresentam desvios padrão maiores, indicando que essas cidades experimentam variações mais significativas nas precipitações de chuva. Além disso, os valores mínimos e máximos registrados mostram que as precipitações podem variar bastante de um mês para outro em todas as cidades. João Pessoa apresenta mínimo de 0 e máximo de 640, enquanto Picuí tem mínimo de 0 e máximo de 370.

Tabela 2 – Estatística Descritiva das Precipitações de chuvas entre 1991-01-01 à 2019-12-01.

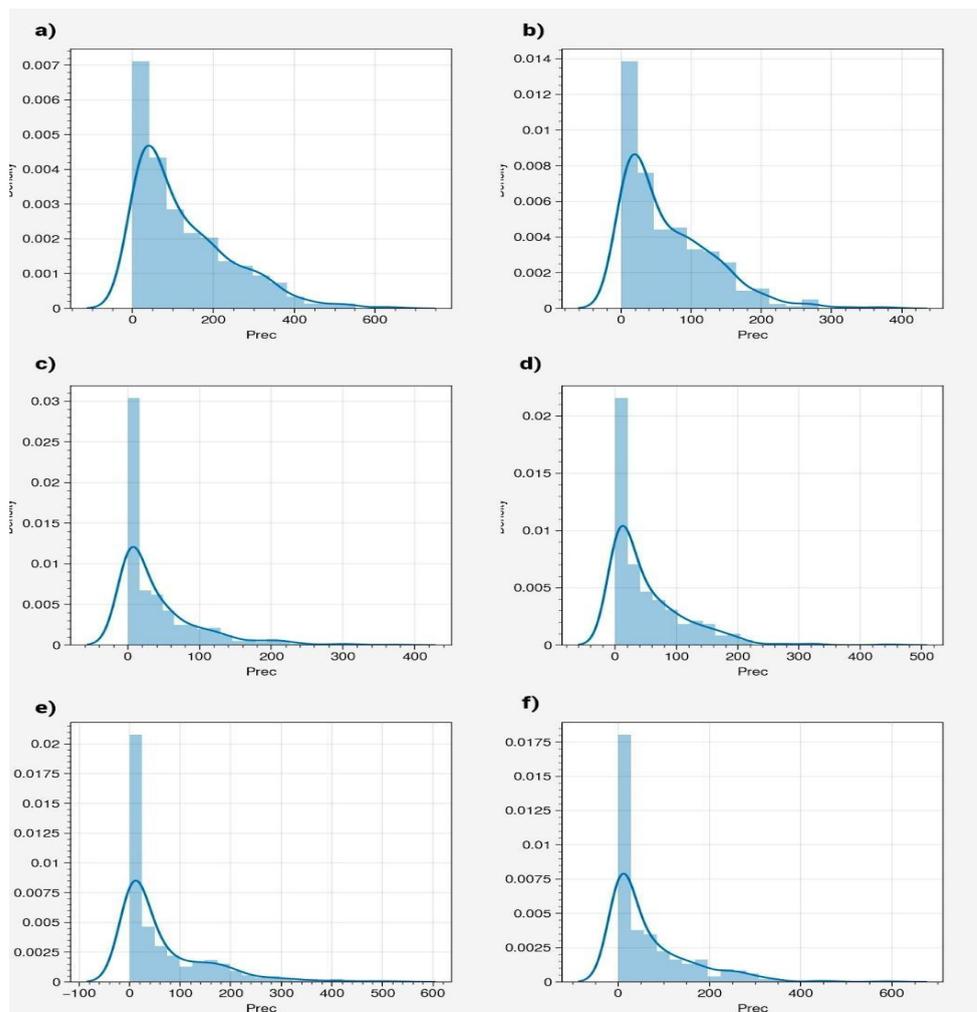
	Latitude	Longitude	\bar{x}	sd	p25	p75	p100
Campina Grande	-7.23072	-35.8817	67	64	15	100	380
João Pessoa	-7.11532	-34.861	130	120	34	190	640
Monteiro	-7.8915	-37.1169	52	63	7.1	79	450
Picuí	-6.54727	-36.3621	43	60	1.1	60	370
Patos	-6.02556	-37.2779	64	88	3.8	94	520
Souza	-6.76782	-38.2099	68	91	3.6	46	590

Fonte: Elaborado pelo autor, 2023.

\bar{x} = Média; **sd** = Desvio padrão; **p25**= Primeiro quartil (25° percentil); **p75**= Terceiro quartil (75° percentil);

Observa-se nos histogramas da figura 1, que eles possuem assimetria à esquerda, indicando que a maioria das observações de chuva estão concentradas em valores menores, enquanto há algumas observações de chuva que são muito mais altas do que a média.

Figura 1 - Histogramas das precipitações de chuvas, A -João Pessoa, B - Campina Grande, C -Picuí, D -Monteiro, E -Patos, F -Souza.



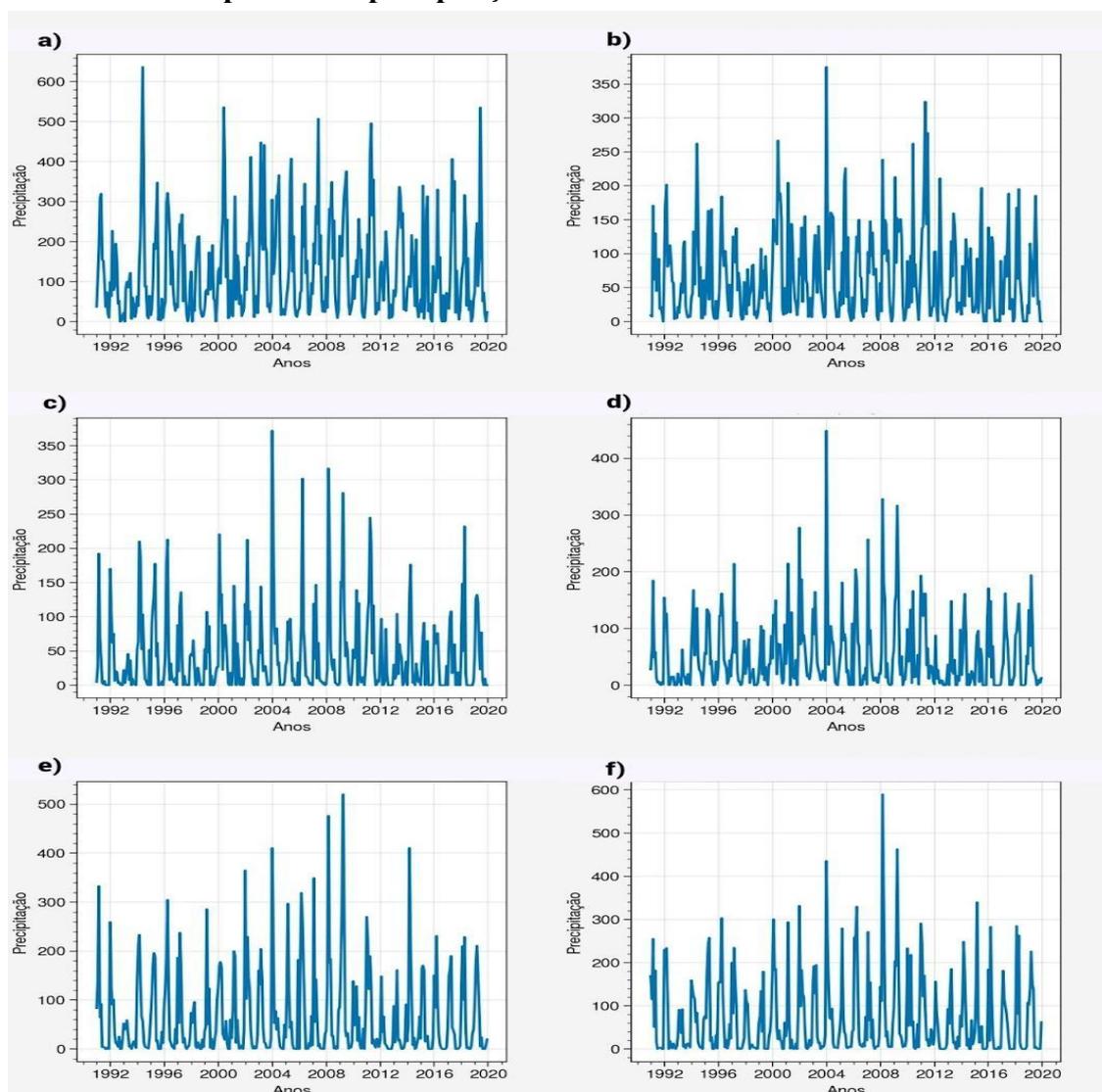
Fonte: Elaborado pelo autor, 2023.

A Figura 2 descreve uma série temporal ilustrando os anos de máxima precipitação pluviométrica nas cidades objeto deste estudo. Em relação a João Pessoa, designada no gráfico como (A), destaca-se que os anos de 1994 e 2019 foram caracterizados por precipitações significativas, atingindo respectivamente 636 mm e 536 mm. Analogamente, para Campina Grande, rotulada como (B), os anos de pico pluviométrico foram 2004 e 2011, com índices de precipitação de 375 mm e 324 mm, respectivamente.

No que concerne a Picuí (C), os anos de 2004 e 2008 emergem como os mais chuvosos, com volumes de precipitação de 371 mm e 316 mm, respectivamente. Em Monteiro (D), os dados indicam que os anos de 2004 e 2007 foram marcados por elevados níveis de precipitação, atingindo 448 mm e 328 mm, respectivamente. Quanto a Patos (E), os anos de 2007 e 2009 registraram os maiores índices pluviométricos, totalizando 476 mm e 520 mm, respectivamente. Finalmente, em Souza (F), os anos de 2008 e 2009 destacaram-se com volumes de precipitação de 589 mm e 462 mm.

De forma geral, esses dados apontam para uma variabilidade temporal significativa no que tange aos índices de precipitação nas cidades analisadas.

Figura 2 - Séries temporais das precipitações de chuvas entre 1991-01-01 à 2019-12-01



Fonte: Elaborado pelo autor, 2023.

3.2 Standardized Precipitation Index (SPI)

Através da aplicação do Índice Padronizado de Precipitação (SPI), os eventos de chuvas e secas foram identificados em diferentes escalas temporais, incluindo períodos de 1 mês, 3 meses, 6 meses, 9 meses, 12 meses e 24 meses. Cada estação de coleta de dados pluviométricos apresenta suas próprias características distintas e variações sazonais, o que resulta em períodos de precipitação abundante ou escassez, além de variações na intensidade e duração desses períodos.

No entanto, mesmo que cada estação apresente variações particulares, eventos climáticos intensos tendem a impactar uma área mais ampla. Quando esses eventos possuem caráter climático, eles podem ser observados abrangendo toda a região de estudo, independente da escala temporal analisada.

Na Figura 3, é retratado o comportamento da média do Índice Padronizado de Precipitação (SPI) com uma escala de um único mês (SPI-1). A análise revela uma notável flutuação mensal, caracterizada por múltiplos picos positivos e negativos de curta duração, bem como alguns de maior extensão. Mesmo dentro desse intervalo de um mês, pode-se notar a presença de sequências de valores positivos em determinados anos nas estações analisadas. Essas sequências indicam a ocorrência de eventos de chuvas que se manifestarão mais claramente em escalas temporais mais abrangentes, fazendo com que seus efeitos sejam percebidos posteriormente e variem de acordo com a duração nas diferentes escalas. Similarmente, no lado negativo da distribuição, podem ser observados períodos de secas de curta duração.

Em relação a João Pessoa, é evidente que os anos de 2003, 2004 e 2013 se destacaram por apresentarem as maiores quantidades de chuva registradas. No entanto, os anos de 2015, 2016 e 2018 foram marcados por secas intensas na mesma região.

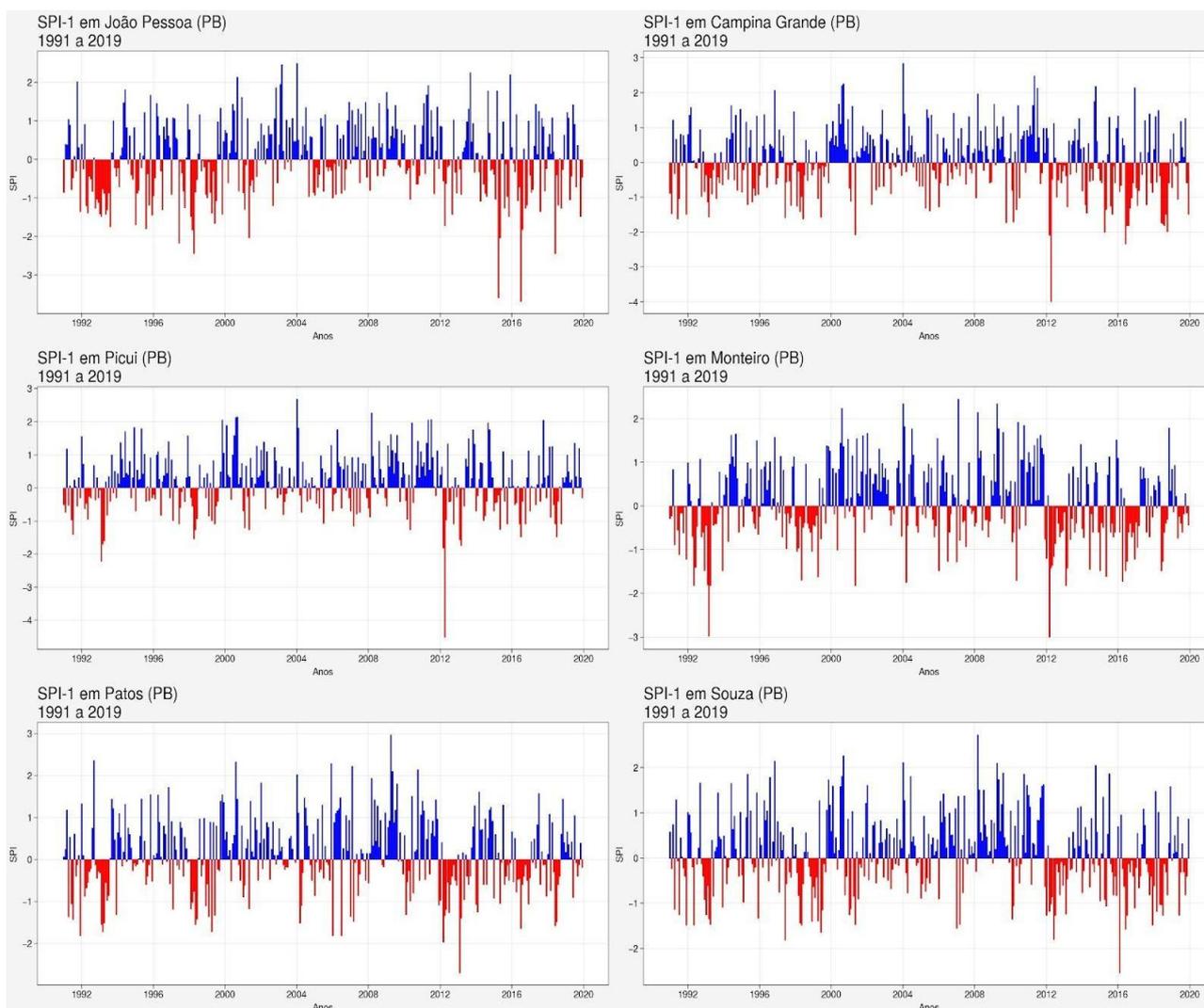
Em Campina Grande, os anos que se sobressaíram pelo maior volume de chuva foram 2000, 2004 e 2011. Contrapondo-se a isso, os anos de 2001, 2012 e 2016 foram caracterizados por secas mais pronunciadas.

Em Picuí, os anos de 2000, 2004 e 2008 se destacaram pelas maiores quantidades de chuva registradas, enquanto os anos de 1993, 2012 e 2013 apresentaram níveis mais acentuados de seca. Analisando a cidade de Monteiro, os anos de 2004, 2007 e 2009 se destacaram por serem períodos mais chuvosos. Em contrapartida, os anos de 1993, 2001 e 2012 foram notáveis por suas secas mais intensas.

No contexto de Patos, os anos de 1992, 2000 e 2009 foram caracterizados por volumes mais significativos de precipitação. Por outro lado, os anos de 1991, 2012 e 2013

foram marcados por períodos de seca mais acentuada. Em Sousa, os anos de 2000, 2008 e 1996 apresentaram um maior volume de chuva em comparação com outros anos. Contudo, os anos de 1997, 2016 e 2017 foram caracterizados por secas mais severas na região.

Figura 3 - SPI 1



Fonte: Elaborado pelo autor, 2023.

Tabela 3 - SPI – 1

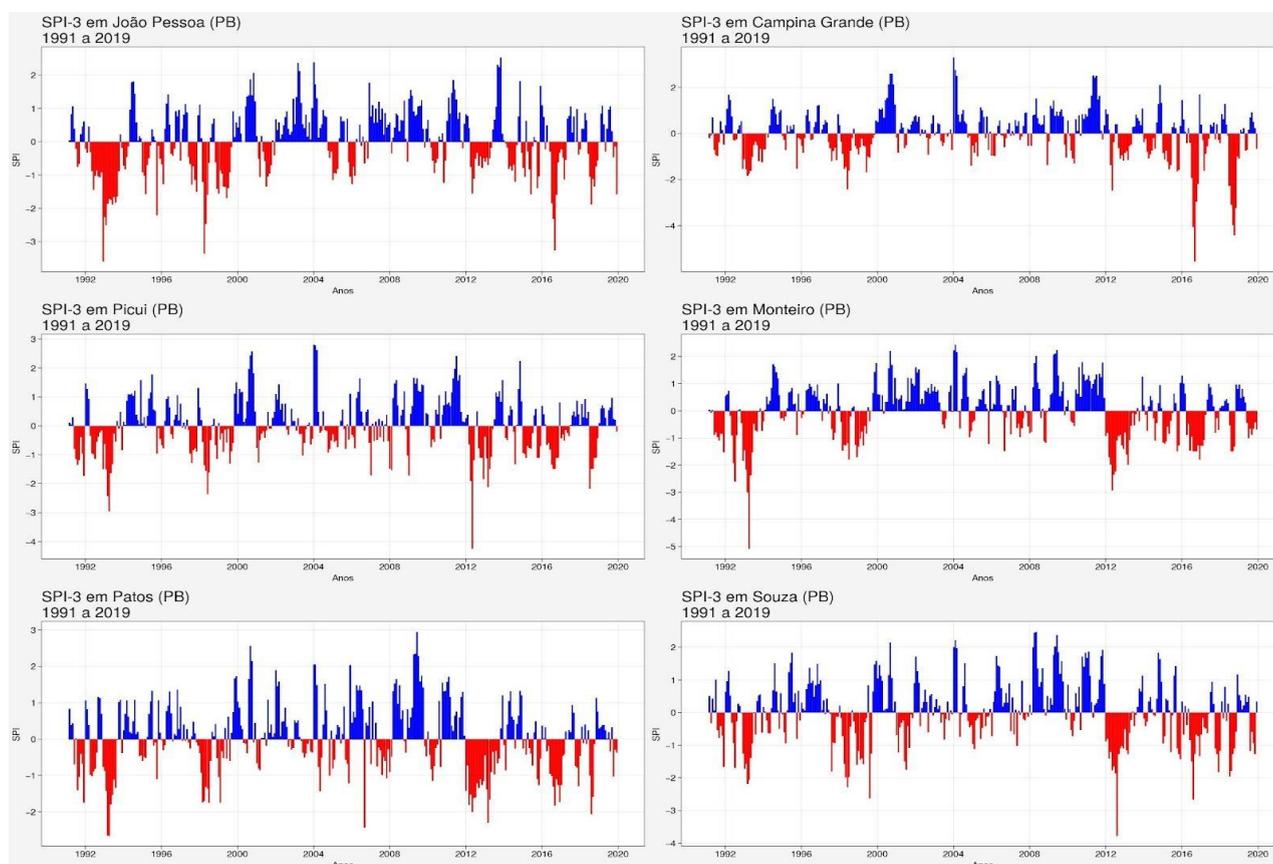
SPI-1	Máximo	Mínimo
Campina Grande	2.853542	-3.994498
João Pessoa	2.498881	-3.685909
Monteiro	2.458819	-3.004865
Picuí	2.695183	-4.514741
Patos	2.975877	-2.707451
Souza	2.733314	-2.542402

Fonte: Elaborado pelo autor, 2023.

Na figura 4, o SPI-3 fornece uma visão mais clara das observações feitas na escala mensal (SPI-1). Foi observado que os eventos mencionados anteriormente persistiram e que a aleatoriedade na escala menor diminuiu. Um ponto importante a ser observado nesta escala de 3 meses é que ela mostrou com mais precisão a intensidade dos eventos, acumulando valores consecutivos de mesmo sinal e ampliando os picos dos eventos mais relevantes.

De acordo com a tabela 4, o valor máximo para o SPI-3 foi registrado em Campina Grande em 2004. Em relação aos eventos negativos, Campina Grande também apresentou o menor valor em 2016.

Figura 4 - SPI 3



Fonte: Elaborado pelo autor, 2023.

Tabela 4 - SPI – 3

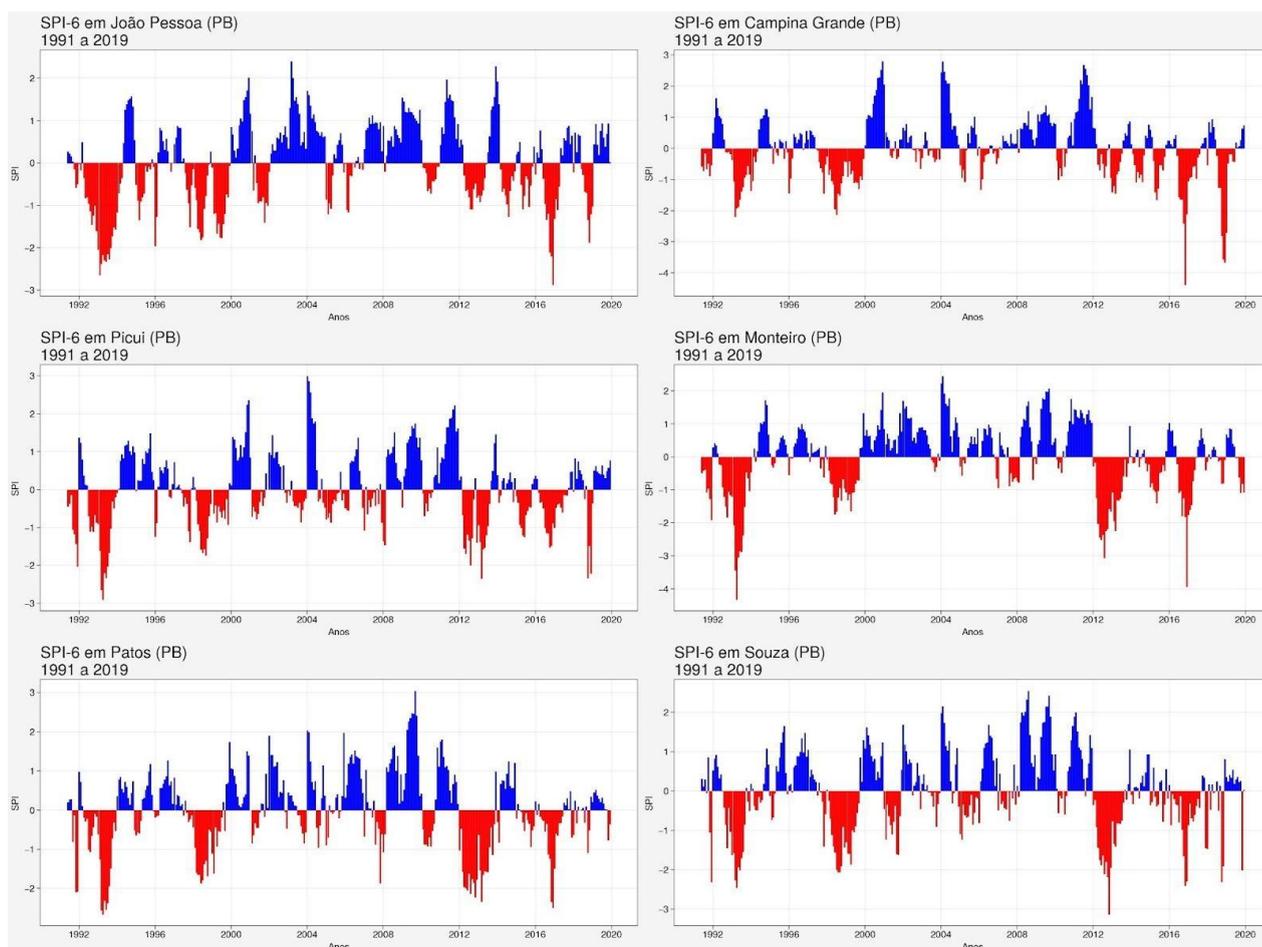
SPI-3	Máximo	Mínimo
Campina Grande	3.316800	-5.553479
João Pessoa	2.539662	-3.585842
Monteiro	2.442202	-5.070721
Picuí	2.816756	-4.243642
Patos	2.952973	-2.658167
Souza	2.478849	-3.776287

Fonte: Elaborado pelo autor, 2023.

Ao utilizar o SPI-6 na figura 5, foi possível identificar a época de transição entre o período chuvoso e o período de seca. Em Monteiro, observa-se um período de chuva que segue em torno de uma média sem grandes picos, ao contrário do período de seca, que apresenta grandes picos predominantes a partir de 1993. Nas estações de Campina Grande e João Pessoa, observam-se longos períodos chuvosos, mais do que nas estações de Patos e Souza. Da mesma forma, no período de seca, as estações de Patos e Souza apresentam maior destaque nesta escala do que as demais, devido à sua localização geográfica.

De acordo com a tabela 5, Campina Grande apresentou o maior valor em 2016 em relação ao evento negativo. Quanto ao SPI-6, o maior valor registrado para o evento positivo foi em Patos em 2009.

Figura 5 - SPI 6



Fonte: Elaborado pelo autor, 2023.

Tabela 5 - SPI – 6

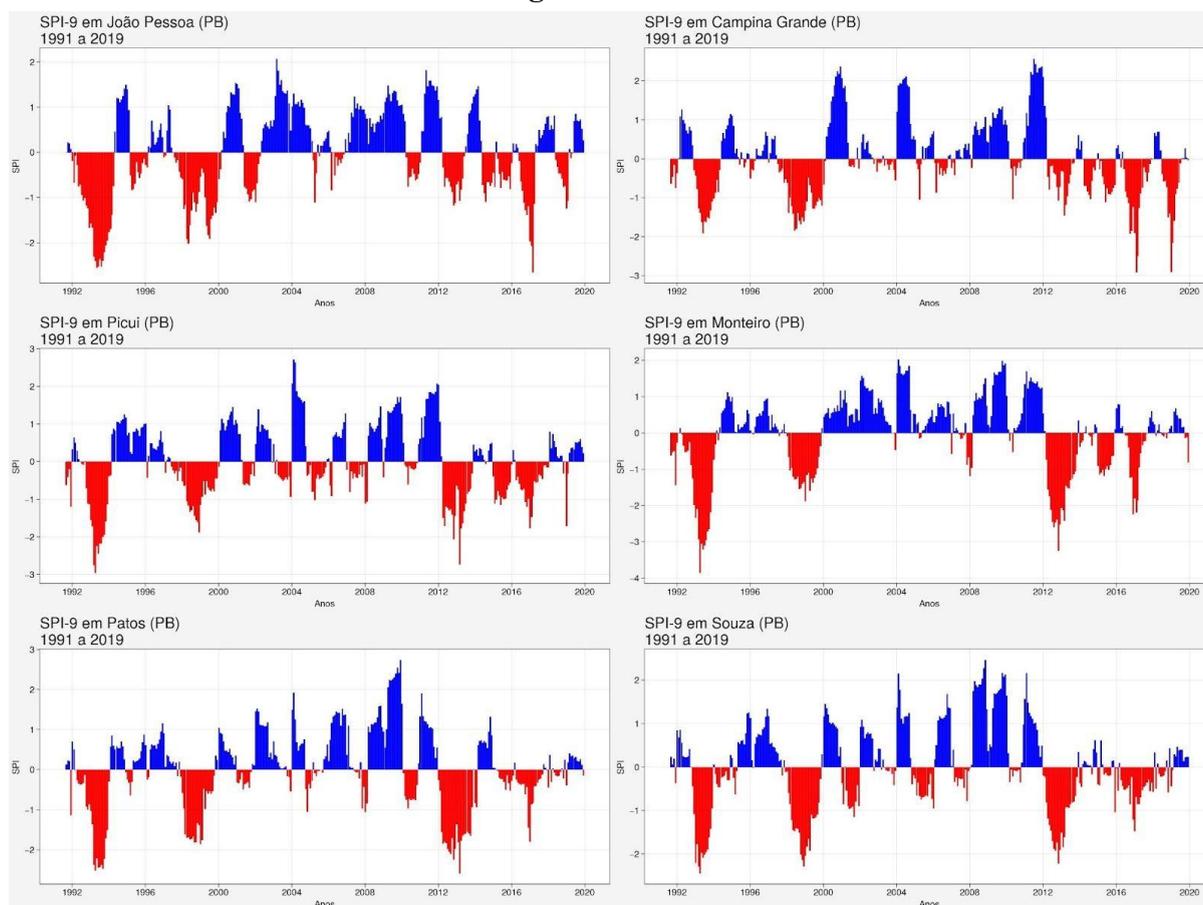
SPI-6	Máximo	Mínimo
Campina Grande	2.805822	-4.382853
João Pessoa	2.407535	-2.875667
Monteiro	2.456994	-4.330417
Picuí	2.998535	-2.902734
Patos	3.046565	-2.664731
Souza	2.549074	-3.134210

Fonte: Elaborado pelo autor, 2023.

Ao utilizar o SPI-9 na figura 6, não foram observadas diferenças significativas em relação à escala de 6 meses (SPI-6) nos períodos de transição das regiões. No entanto, esta escala proporcionou uma melhor definição da oscilação entre períodos positivos e negativos, mantendo os eventos críticos mencionados anteriormente.

De acordo com a tabela 6, para os registros de eventos críticos no SPI-9, Monteiro obteve o maior valor de precipitação negativa, enquanto Patos apresentou o maior evento positivo. O menor valor, ou seja, o maior pico de seca registrado foi em Palmares no ano de 2017.

Figura 6 - SPI 9



Fonte: Elaborado pelo autor, 2023.

Tabela 6 - SPI – 9

SPI-9	Máximo	Mínimo
Campina Grande	2.569257	-2.909613
João Pessoa	2.076713	-2.656554
Monteiro	2.026951	-3.850753
Picuí	2.725501	-2.955701
Patos	2.745890	-2.587995
Souza	2.465498	-2.441443

Fonte: Elaborado pelo autor, 2023.

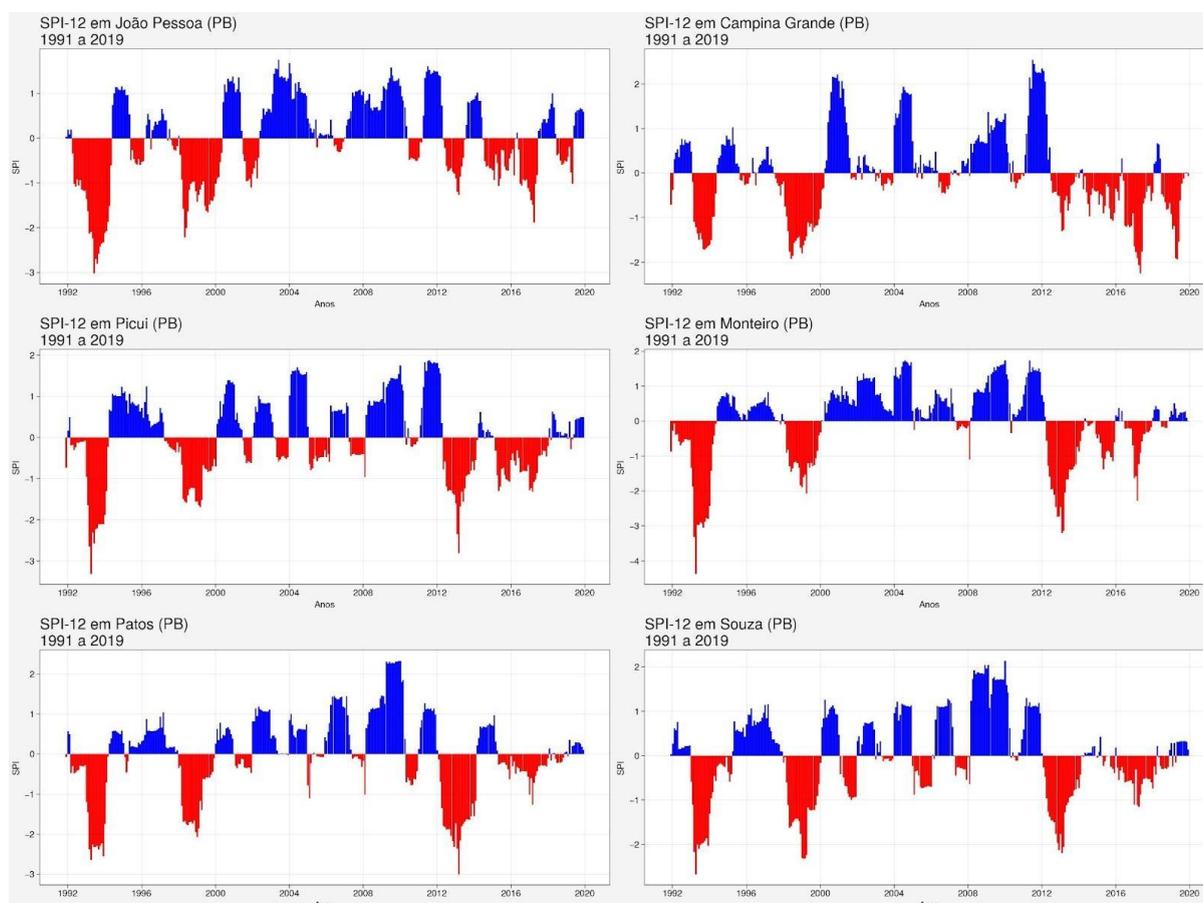
Ao analisar a escala anual registrada pelo SPI-12 na figura 7 para cada estação do estado, podemos observar os seguintes padrões climáticos. Em João Pessoa, foram registrados longos períodos chuvosos em 2003 e 2008, com exceção de um período seco significativo que ocorreu entre 1992, 1998 e 2015. Em Campina Grande, houve uma grande variabilidade com picos positivos e negativos, mas a partir da década de 2011, as secas se tornaram mais frequentes. Em Monteiro, destacam-se os longos períodos chuvosos entre

2000 e 2012, enquanto as secas ocorreram principalmente entre 1992 e 1994, e prevaleceram no período de 2012 a 2017. Picuí apresenta uma grande variabilidade nas chuvas, com destaque para os anos de 1995 a 1997 e 2008 a 2009, e períodos de seca entre 1997 e 1999, e 2012 a 2017.

Nos municípios de Sousa e Patos, localizados no sertão do estado, também foram observadas variações nos períodos de chuva. Em Sousa, destacam-se os períodos de seca em 1997 e 2012, enquanto em Patos, a seca passou a ser predominante a partir de 2012.

A tabela 7 apresenta os valores máximos e mínimos registrados para cada estação no SPI-12, localizado no Agreste Paraibano. Em Campina Grande, foi registrado o maior evento positivo em 2011, enquanto Monteiro teve o menor evento em comparação com as outras estações. Em relação aos eventos negativos, relacionados à seca, o valor mínimo foi registrado em Monteiro no ano de 1993.

Figura 7 - SPI 12



Fonte: Elaborado pelo autor, 2023.

Tabela 7 - SPI – 12

SPI-12	Máximo	Mínimo
Campina Grande	2.546700	-2.246363
João Pessoa	1.751131	-3.014405
Monteiro	1.743142	-4.366061
Picuí	1.876263	-3.310416
Patos	2.329988	-2.999056
Souza	2.140688	-2.673100

Fonte: Elaborado pelo autor, 2023.

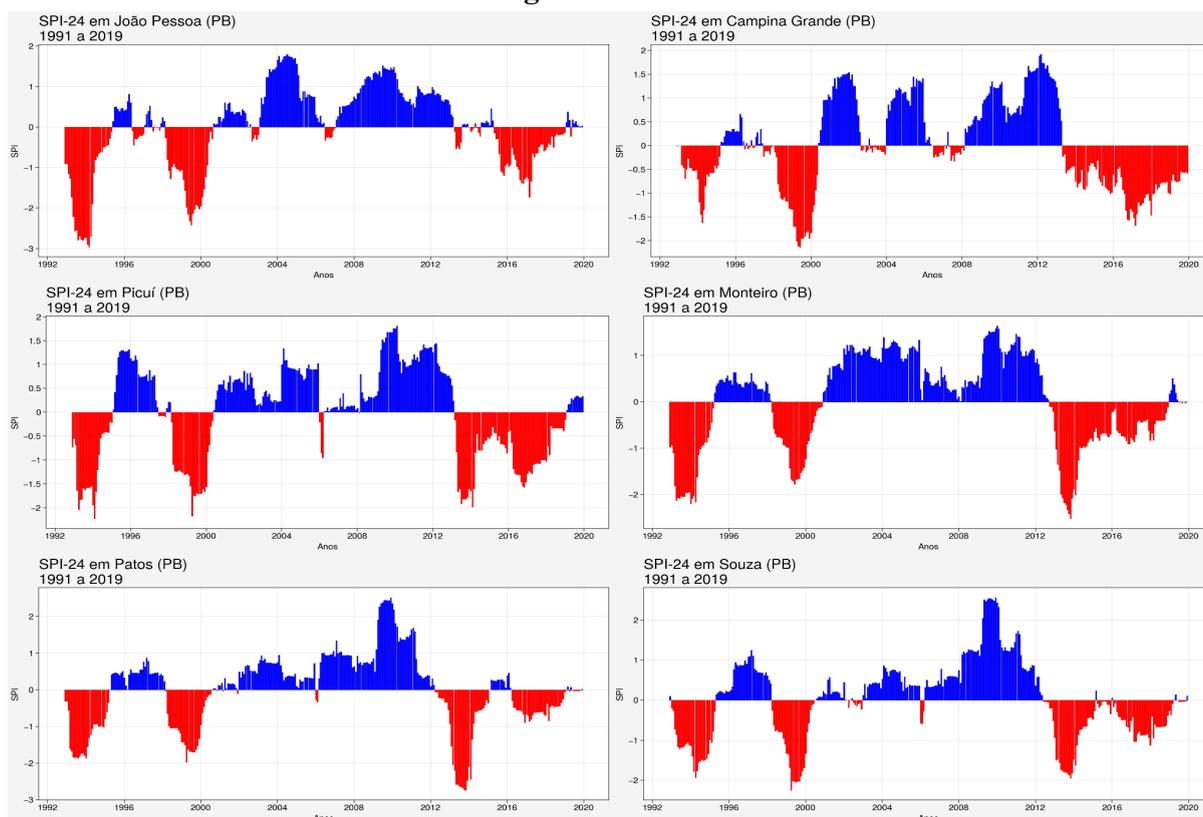
A métrica SPI-24 é um índice de Precipitação Padronizada calculado para uma janela temporal de 24 meses, fornecendo assim uma perspectiva abrangente das condições hídricas—sejam elas de umidade ou de seca—em uma localidade específica. Conforme ilustrado na Figura 8, observa-se uma variação significativa nas condições de umidade em seis estações meteorológicas selecionadas durante o intervalo de 2000 a 2012, caracterizado predominantemente por períodos de elevada umidade. Contrapondo-se a esse intervalo, os dados coletados entre os anos de 2013 e 2019 revelam uma tendência de intensificação de condições de seca.

Na Tabela 8, o intervalo de variação do SPI-24 nas diferentes estações meteorológicas aponta para flutuações significativas nas condições hídricas. Por exemplo, em Campina Grande, o valor máximo de SPI-24 foi registrado como 1.924587, enquanto o mínimo foi -2.138502. Essa amplitude sugere que essa região experimentou tanto períodos de umidade significativa quanto de seca intensa ao longo do período de análise.

Similarmente, João Pessoa e Monteiro apresentam valores máximos e mínimos que indicam a ocorrência de condições tanto de seca quanto de umidade. O intervalo entre os valores máximo e mínimo é especialmente amplo em João Pessoa, onde o SPI-24 varia de 1.800698 a -2.963429, enfatizando a variabilidade extrema das condições hídricas na localidade.

É importante destacar também as localidades de Patos e Souza, onde foram registrados os valores máximos mais elevados de SPI-24, sendo 2.518448 e 2.563850, respectivamente. Estes valores podem indicar episódios de umidade excepcionalmente elevada.

Figura 8 - SPI 24



Fonte: Elaborado pelo autor, 2023.

Tabela 8 - SPI – 24

SPI-12	Máximo	Mínimo
Campina Grande	1.924587	-2.138502
João Pessoa	1.800698	-2.963429
Monteiro	1.641876	-2.522586
Picuí	1.812801	-2.229101
Patos	2.518448	-2.742311
Souza	2.563850	-2.249191

Fonte: Elaborado pelo autor, 2023.

3.3 Modelagem

Analisando os resultados na tabela 8, obtidos pelos modelos de Machine Learning aplicados - Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR) e Random Forest Regressor (RF) - nas seis estações (Campina Grande, João Pessoa, Monteiro, Picuí, Patos e Souza), é possível discorrer sobre as métricas de desempenho (MAE, MSE, RMSE, R^2 , RMSLE e MAPE).

Em Campina Grande o Extra Trees Regressor destacou-se com um R^2 de 0.7515, indicando que o modelo foi capaz de explicar 75.15% da variância nos dados de precipitação

SPI 9. Este modelo também apresentou o menor valor para as métricas de erro MAE, MSE e RMSE em comparação com os outros modelos para essa estação. Random Forest Regressor, por outro lado, apresentou uma performance um pouco inferior ao ET, mas ainda relativamente sólida, com um R^2 de 0.5972. Gradient Boosting Regressor foi o modelo com o menor desempenho nesta estação, revelando o maior MAE, MSE e RMSE, e o menor valor de R^2 (0.4489) entre os três modelos.

Em João Pessoa Extra Trees Regressor obteve o melhor desempenho entre os modelos com um R^2 de 0.7722, sendo o modelo que melhor explica a variabilidade dos dados em João Pessoa. Random Forest Regressor e Gradient Boosting Regressor demonstraram desempenhos semelhantes em termos de R^2 , porém, o GBR resultou em um RMSE e MAE ligeiramente superior ao rf, indicando que pode ter havido uma maior dispersão dos erros.

Analisando de forma mais abrangente, o ET consistentemente produziu o maior valor de R^2 em todas as estações, indicando uma alta explicabilidade dos dados.

O Random Forest Regressor exibiu uma consistência em suas previsões, mantendo valores razoavelmente estáveis para o R^2 através de todas as estações.

O Gradient Boosting Regressor mostrou-se o mais variável entre os modelos, com uma diferença notável em seu desempenho entre as diferentes estações. Isso pode ser um indicativo de sensibilidade a certas características dos dados ou uma sinalização de que os parâmetros do modelo possam necessitar de um ajuste mais refinado.

O Erro Absoluto Médio e Erro Quadrático Médio: Para todas as estações, o Extra Trees Regressor demonstrou, de forma geral, valores inferiores para MAE e MSE em comparação aos outros modelos, o que indica uma melhor precisão e confiabilidade nos dados previstos. A Raiz do Erro Quadrático Médio: Em termos de RMSE, o ET também se destacou, pois valores mais baixos de RMSE indicam melhor desempenho do modelo.

O Erro Quadrático Médio Logarítmico: Esta métrica é particularmente interessante para a modelagem de previsão, onde a escala do erro é uma consideração importante, sendo essencial observar as tendências através dos diferentes modelos e estações.

Tabela 9 - Comparativo de Desempenho dos Modelos de Aprendizado de Máquina Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR) e Random Forest Regressor (RF) Aplicados em Seis Estações Meteorológicas: Campina Grande, João Pessoa, Monteiro, Picuí, Patos e Sousa para o SPI-9.

Município	Modelo	MAE	MSE	RMSE	R ²	RMSLE	MAPE
Campina Grande	et	0.3011	0.1963	0.4291	0.7515	0.2022	2.2065
	rf	0.3702	0.3052	0.5392	0.5972	0.2487	3.7448
	gbr	0.4404	0.3626	0.5899	0.4489	0.2804	5.3000
João Pessoa	et	0.3024	0.2092	0.4373	0.7722	0.2076	0.9896
	rf	0.3852	0.3147	0.5375	0.6524	0.2469	1.7170
	gbr	0.4641	0.3961	0.6133	0.5566	0.2924	2.6099
Monteiro	et	0.2706	0.1629	0.3939	0.8440	0.1864	1.6337
	rf	0.3102	0.2300	0.4705	0.7656	0.2181	1.9983
	gbr	0.3554	0.2683	0.5120	0.7174	0.2314	2.6924
Picuí	et	0.3214	0.2194	0.4408	0.7763	0.2176	0.9097
	rf	0.3343	0.2724	0.4837	0.7257	0.2160	0.9919
	gbr	0.4018	0.3303	0.5475	0.6527	0.2573	1.2020
Patos	et	0.2772	0.1676	0.3881	0.8377	0.1953	1.0255
	rf	0.3221	0.2225	0.4532	0.7774	0.2238	1.5633
	gbr	0.3761	0.2694	0.5077	0.7211	0.2474	1.4474
Souza	et	0.2980	0.1605	0.3979	0.8170	0.2072	2.1041
	rf	0.3060	0.2029	0.4413	0.7644	0.2229	1.4333
	gbr	0.3672	0.2772	0.5149	0.6792	0.2601	1.848

Fonte: Elaborado pelo autor, 2023.Sg

Observa-se na tabela 10, que em Campina Grande, o Extra Trees Regressor demonstrou um R² robusto de 0.8442, destacando-se como o modelo que mais explicou a variabilidade dos dados. Ele também mostrou solidez em outras métricas, apresentando os menores valores de MAE, MSE, e RMSE, o que reflete sua habilidade em minimizar os erros nas previsões. Por outro lado, o Gradient Boosting Regressor registrou maior sensibilidade e erros, principalmente destacado por um RMSLE de 0.2678 e um MAPE de 2.0509, indicando que suas previsões podem ter sido desafiadas por valores atípicos ou uma distribuição assimétrica dos resíduos.

João Pessoa, o ET novamente apresenta a melhor performance em termos de R² (0.8631), contudo, é vital notar uma variação considerável nas métricas de erro entre os modelos, com o gbr exibindo uma tendência de previsões menos precisas, indicada por seus maiores valores de MAE e MSE, por exemplo. Mesmo com o maior R², o ET apresentou um

MAPE de 1.0797, implicando que o modelo pode ainda ser susceptível a erros sistemáticos e que uma avaliação adicional dos resíduos pode ser construtiva.

Em Monteiro, os modelos revelam métricas de erro mais próximas entre si, ainda que o ET prevaleça em termos de ajuste global ($R^2=0.8895$). Aqui, o GBR demonstrou uma melhora relativa em termos de consistência, embora ainda apresente o maior RMSLE e MAPE entre os modelos.

Em Picuí, o modelo ET se sobressai com um coeficiente de determinação (R^2) de 0.8143, oferecendo um adequado ajuste ao padrão dos dados. Ainda assim, quando observamos o MAPE elevado (1.2998), que indica uma certa amplitude nos erros de previsão em termos percentuais, surgem questionamentos sobre a distribuição do erro ao longo do período analisado. O GBR revela a maior volatilidade nas métricas, sugerindo que pode estar se debatendo com a presença de ruído ou padrões menos óbvios nos dados, demandando, talvez, uma sintonia mais fina dos parâmetros.

Para Patos, as métricas retratam uma narrativa onde o ET domina em termos de métricas de erro e ajuste (R^2 de 0.8543). O GBR, apresentando o menor R^2 e os maiores valores para o MAE e MSE, indica que suas previsões foram, em geral, menos precisas e consistentes. Com um MAPE de 1.4735, o ET ainda sugere que possíveis melhorias podem ser exploradas, talvez através da integração de variáveis adicionais ou ajustes no tratamento de outliers.

Souza mostra-se um caso interessante. Embora o ET apresente um alto $R^2(0.8609)$, um olhar para o MAPE revela uma disparidade notável (4.5121). Esta divergência entre um R^2 robusto e um MAPE elevado pode indicar que, enquanto o modelo captura a variabilidade geral dos dados, ele pode estar lutando com a previsão de certos pontos específicos – possivelmente aqueles com altos valores de SPI, que teriam um impacto mais acentuado sobre o MAPE. O Random Forest Regressor, apesar de exibir um desempenho geral sólido, também exibe um MAPE elevado (6.1354), indicando que erros sistemáticos podem estar em jogo e merecem uma investigação mais minuciosa.

Tabela 10 - Comparativo de Desempenho dos Modelos de Aprendizado de Máquina Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR) e Random Forest Regressor (RF) Aplicados em Seis Estações Meteorológicas: Campina Grande, João Pessoa, Monteiro, Picuí, Patos e Sousa para o SPI-12.

Município	Modelo	MAE	MSE	RMSE	R ²	RMSLE	MAPE
Campina Grande	et	0.2733	0.1367	0.3666	0.8442	0.1898	1.0606
	rf	0.3657	0.2789	0.5202	0.6612	0.2308	1.6131
	gbr	0.4329	0.3357	0.5733	0.6133	0.2678	2.0509
João Pessoa	et	0.2575	0.1263	0.3479	0.8631	0.1900	1.0797
	rf	0.3995	0.2910	0.5301	0.6846	0.2649	1.3565
	gbr	0.5132	0.4183	0.6411	0.5397	0.3342	1.5097
Monteiro	et	0.2294	0.1346	0.3482	0.8895	0.1596	12.5304
	rf	0.2862	0.2282	0.4605	0.8055	0.2142	15.9226
	gbr	0.3452	0.2531	0.4838	0.7867	0.2317	16.1387
Picuí	et	0.3030	0.1874	0.4179	0.8143	0.2053	1.2998
	rf	0.3462	0.2485	0.4851	0.7421	0.2269	1.8960
	gbr	0.4627	0.3627	0.5852	0.6342	0.2766	2.4618
Patos	et	0.2671	0.1509	0.3799	0.8543	0.1966	1.4735
	rf	0.2885	0.2012	0.4388	0.7993	0.2070	1.9101
	gbr	0.3780	0.2860	0.5256	0.7150	0.2554	2.6850
Sousa	et	0.2583	0.1371	0.3504	0.8609	0.1743	4.5121
	rf	0.2901	0.1983	0.4338	0.7798	0.2090	6.1354
	gbr	0.3640	0.2432	0.4827	0.7345	0.2385	6.2883

Fonte: Elaborado pelo autor, 2023.

De acordo com a tabela 11, os modelos para Campina Grande no contexto de SPI 24 refletem uma alta capacidade de explicar a variância nos dados, sendo o Extra Trees Regressor novamente o modelo líder com um R² de 0.9542. Contudo, a interpretação precisa ser cuidadosa ao observar o MAPE, que é relativamente alto (5.8258) e aponta para uma significativa discrepância percentual nas previsões deste modelo. Uma investigação sobre a distribuição dos erros ao longo do tempo e em diferentes magnitudes de SPI seria uma etapa crucial para entender a origem desta discrepância.

João Pessoa segue uma narrativa semelhante, com o ET ostentando um alto R² (0.9517), mas igualmente com um MAPE contrastante (0.9092). Nota-se que os modelos conseguem, de forma geral, capturar os padrões nas séries temporais, mas este alto MAPE pode indicar episódios de erros mais expressivos que merecem atenção. O GBR, enquanto captura uma proporção razoável da variabilidade dos dados (R² de 0.8269), sua métrica

MAPE sugere que a precisão de suas previsões pode ser problemática em certos períodos ou faixas de SPI.

Em Monteiro, o ET exibe um robusto R^2 de 0.9726, mostrando forte adequação às variações nos dados de SPI 24. As outras métricas como MAE, MSE e RMSE também são as menores entre os modelos, indicando a solidiedade desse modelo nas previsões..

Picuí apresenta um quadro onde o ET mantém uma boa explicabilidade (R^2 de 0.9374). O MAPE do GBR destaca-se por ser o maior entre os modelos (0.7079), apontando para uma discrepância mais acentuada em suas previsões e valendo uma investigação adicional sobre como seus parâmetros estão sendo influenciados pelos padrões nos dados e se uma recalibração é necessária.

Para Patos, os modelos de maneira geral exibem um desempenho respeitável, porém o contraste entre o R^2 e o MAPE no ET (0.9397 e 0.8697, respectivamente) sugere que os erros percentuais nas previsões não são uniformes e podem ser influenciados por padrões específicos ou outliers nos dados que valem uma exploração mais profunda.

Finalmente, em Souza, enquanto os modelos possuem uma capacidade notável de mapear a variabilidade dos dados (com ET liderando com R^2 de 0.9401), o MAPE sugere que refinamentos adicionais na modelagem ou na seleção de variáveis poderiam prover melhorias, minimizando os erros percentuais nas previsões.

Tabela 11 - Comparativo de Desempenho dos Modelos de Aprendizado de Máquina Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR) e Random Forest Regressor (RF) Aplicados em Seis Estações Meteorológicas: Campina Grande, João Pessoa, Monteiro, Picuí, Patos e Sousa para o SPI-24.

Município	Modelo	MAE	MSE	RMSE	R ²	RMSLE	MAPE
Campina Grande	et	0.1436	0.0389	0.1920	0.9542	0.1037	5.8258
	rf	0.1958	0.0716	0.2630	0.9165	0.1438	5.8953
	gbr	0.2768	0.1401	0.3631	0.8375	0.1871	7.3998
João Pessoa	et	0.1437	0.0485	0.2059	0.9517	0.1086	0.9092
	rf	0.2313	0.1144	0.3325	0.8832	0.1612	2.6106
	gbr	0.3131	0.1756	0.4118	0.8269	0.1966	2.0440
Monteiro	et	0.1077	0.0261	0.1565	0.9726	0.0873	0.4166
	rf	0.1794	0.0672	0.2579	0.9306	0.1256	0.4778
	gbr	0.2402	0.1104	0.3286	0.8827	0.1663	0.9735
Picuí	et	0.1444	0.0546	0.2256	0.9374	0.1217	0.5456
	rf	0.2167	0.1049	0.3138	0.8788	0.1581	0.5177
	gbr	0.2941	0.1798	0.4180	0.7907	0.2051	0.7079
Patos	et	0.1437	0.0536	0.2253	0.9397	0.1158	0.8697
	rf	0.2010	0.0974	0.3071	0.8909	0.1493	1.0662
	gbr	0.2882	0.1698	0.4093	0.8130	0.2023	1.7234
Sousa	et	0.1431	0.0482	0.2111	0.9401	0.1173	1.2024
	rf	0.1933	0.0909	0.2905	0.8914	0.1438	1.4080
	gbr	0.2678	0.1383	0.3656	0.8308	0.1863	1.3774

Fonte: Elaborado pelo autor, 2023.

3.4 Previsões

Explorando com meticulosidade as previsões relacionadas ao Índice Padronizado de Precipitação (SPI) em diferentes localidades da Paraíba, mais especificamente em Campina Grande, João Pessoa, Monteiro, Picuí, Patos e Sousa, visando um profundo entendimento acerca das projeções climáticas e suas implicações nos anos de 2020 e 2021. O SPI, enquanto indicador amplamente utilizado para caracterizar a variação na precipitação e identificar eventos de seca e excesso de chuvas, nos oferece uma visão sistemática e esclarecedora sobre os potenciais cenários de variabilidade e mudança climática nas regiões analisadas. Neste estudo, optamos pela implementação do modelo Extra Trees (ET) para a previsão do SPI em diferentes estações e localidades, dado que demonstrou um melhor ajuste aos dados e uma notável capacidade preditiva em cenários de complexidade climática, foram aplicadas as previsões para os anos de 2020 e 2021.

Observa-se no gráfico apresentado na Figura 9, a previsão do SPI 9 para Campina Grande mostra um padrão de seca persistente para os anos de 2020 e 2021, com valores de SPI predominantemente negativos. Essa tendência indica períodos de seca, onde valores abaixo de -2 são especialmente preocupantes, representando seca extrema. Essa situação pode implicar em estresse hídrico, afetando os recursos hídricos e potencialmente a produção agrícola da região.

Em contraste, João Pessoa exibe uma condição relativamente mais úmida ao longo dos dois anos, com valores do SPI na maior parte do tempo positivos ou próximos de zero. A tendência de hidrologia aqui é mais positiva, com valores indicando normalidade ou condições mais úmidas, sendo benéficas para a agricultura e outros usos dos recursos hídricos.

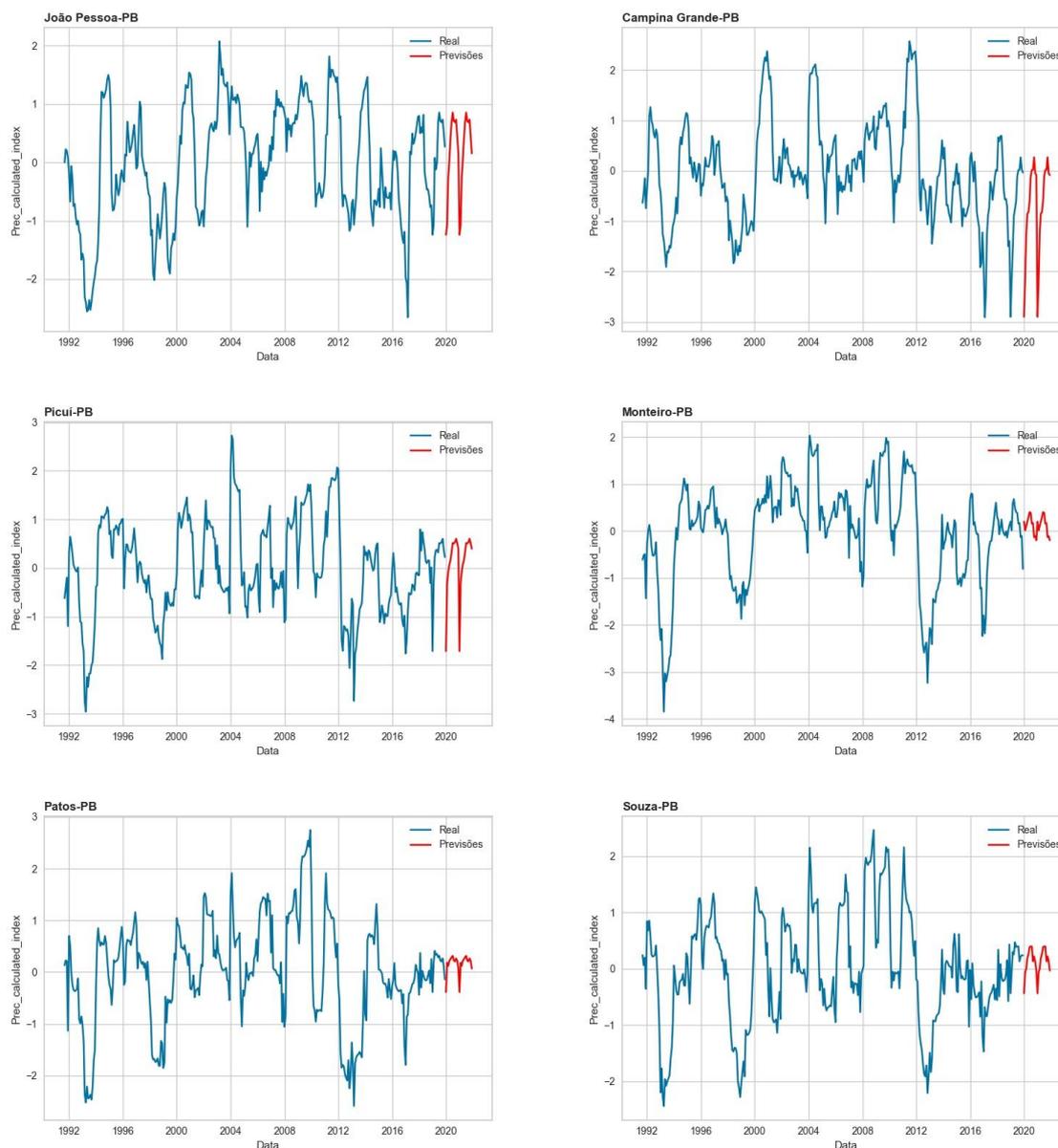
Monteiro mostra uma situação mista, com SPI variando ao redor de zero, sugerindo condições flutuantes entre períodos secos e úmidos. Vale a pena notar a tendência de seca leve a moderada nos finais dos anos, o que pode ser um sinal de padrão sazonal que necessita de mais investigação.

A cidade de Picuí apresenta um padrão interessante, começando o ano com uma seca significativa mas melhorando consistentemente ao longo do ano. Aqui, os meses iniciais de cada ano podem ser críticos para os agricultores e gestores de recursos hídricos, e estratégias de mitigação para estes períodos podem ser necessárias.

Patos apresenta uma tendência predominantemente positiva no SPI, sugerindo condições úmidas ou de normalidade ao longo dos dois anos. No entanto, o final de 2021 sugere uma tendência para condições mais secas, que pode ser um indicador para potenciais problemas no futuro se essa tendência continuar.

Em Souza, há uma oscilação notável nas previsões do SPI, alternando entre condições secas e úmidas. Este padrão oscilante pode indicar uma variabilidade hídrica significativa na região, o que pode apresentar desafios em termos de gestão de recursos hídricos e planejamento agrícola.

Figura 9 - Gráficos de linha das 6 estações referente às previsões do SPI 9



Fonte: Elaborado pelo autor, 2023.

As previsões do Índice de Precipitação Padronizado (SPI) para um período de 12 meses em diversas localidades oferecem uma perspectiva detalhada sobre as tendências de precipitação e, conseqüentemente, sobre potenciais cenários de seca ou excesso de precipitação. De acordo com os gráficos da figura 9, temos os seguintes resultados:

Campina Grande demonstra, ao longo dos anos de 2020 e 2021, uma tendência de SPI 12 majoritariamente negativo, indicando condições de seca. Os valores apresentados, especialmente no primeiro semestre de ambos os anos, sinalizam uma situação de seca

moderada a severa, requerendo estratégias robustas para a gestão de recursos hídricos e adoção de medidas mitigadoras.

A cidade de João Pessoa apresenta uma dinâmica diferenciada, com uma tendência predominantemente positiva do SPI 12. Isto sugere uma condição geral mais úmida e, conseqüentemente, menor susceptibilidade a períodos de seca, ao contrário, possíveis atenções podem ser voltadas para gestões de períodos mais chuvosos.

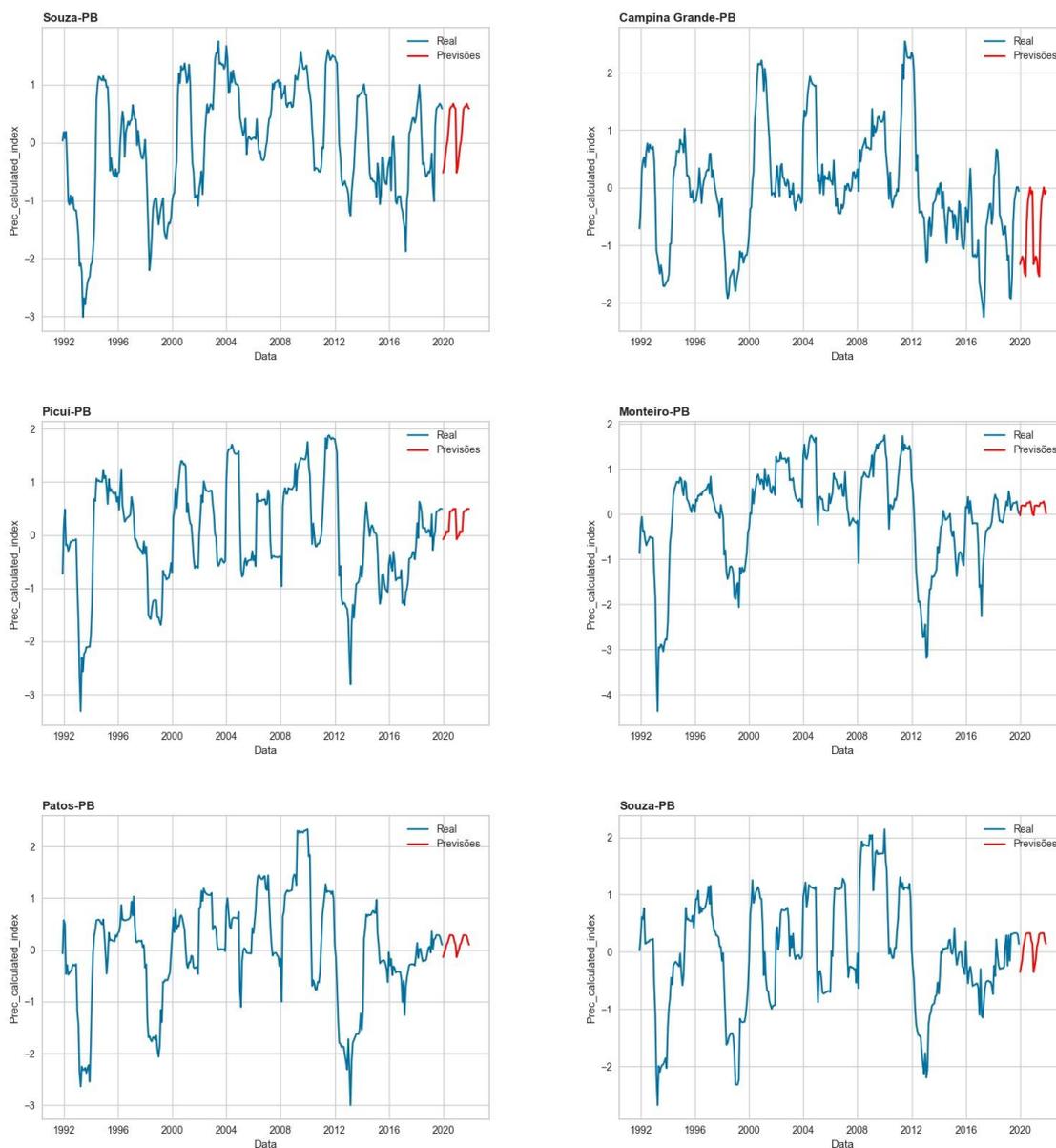
Monteiro revela uma consistência no SPI 12 predominantemente positiva, embora com magnitudes menores comparativamente a João Pessoa. A estabilidade nos índices sugere uma relativa segurança hídrica, no entanto, a vigilância constante é crucial para garantir a sustentabilidade dos recursos.

Similarmente, Picuí exhibe uma tendência positiva do SPI 12, indicando condições mais úmidas, especialmente no segundo semestre dos anos analisados. Isso propõe um cenário hídrico estável, mas a gestão de recursos deve ser aprimorada para evitar potenciais desafios associados ao excesso de precipitação.

Patos mostra uma evolução positiva gradual no SPI 12 ao longo dos meses, sugerindo uma transição para condições mais úmidas. Contudo, a alternância e proximidade de índices negativos indicam a necessidade de estratégias adaptativas para lidar com variabilidades na disponibilidade de água.

A cidade de Souza revela um padrão onde o SPI 12 aumenta progressivamente ao longo do ano, indicando uma recuperação das condições hídricas após um início de ano mais seco. Este perfil reforça a importância de estratégias de gestão de água que sejam resilientes a variações sazonais.

Figura 10 - Gráficos de linha das 6 estações referente às previsões do SPI 12



Fonte: Elaborado pelo autor, 2023.

Nesta seção, faremos uma análise das previsões do SPI-24, como observados nos gráficos da Figura 10, para as diferentes estações meteorológicas, proporcionando uma análise comparativa e insights valiosos sobre padrões climáticos futuros.

A estação de Campina Grande, durante os anos de 2020 e 2021, apresenta consistentemente valores negativos de SP-I24, indicando períodos de seca. A intensidade da seca varia, no entanto, a presença persistente desse padrão sugere condições atmosféricas relativamente secas e implica possíveis desafios em atividades dependentes de água e manejo de recursos hídricos na região.

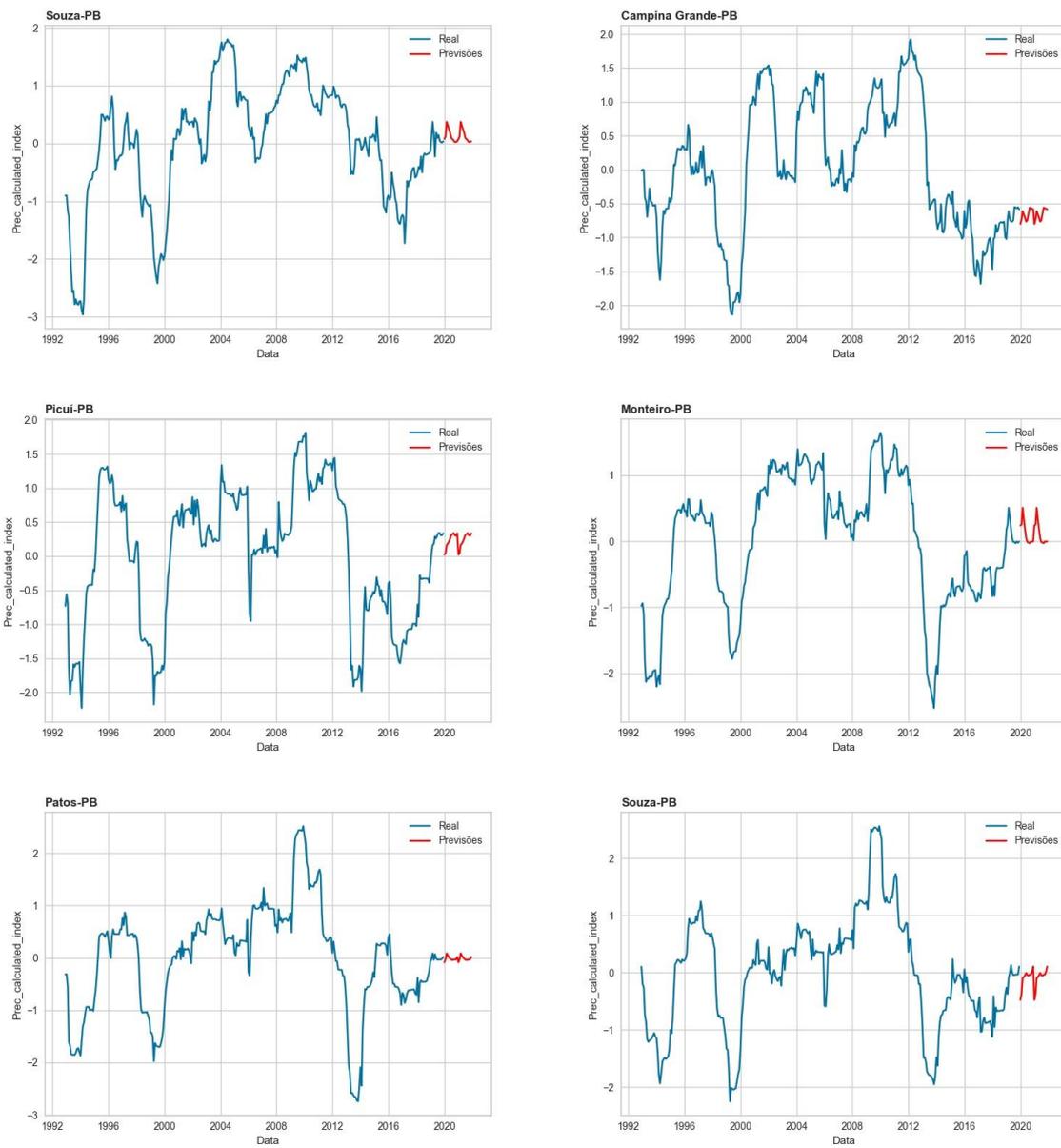
Os dados para João Pessoa indicam uma condição relativamente estável e úmida ao longo dos anos observados, com valores de SPI-24 oscilando majoritariamente em território positivo

Monteiro apresenta uma trajetória interessante, iniciando os anos com índices positivos de SPI-24, indicando períodos mais úmidos, e gradualmente regredindo para valores negativos à medida que os anos avançam.

Picuí mostra um padrão generalizadamente positivo no SPI-24 ao longo dos dois anos, sugerindo uma tendência para condições mais úmidas.

Patos retrata uma situação mista, oscilando entre valores negativos e positivos de SPI-24 ao longo dos anos. Essa alternância entre períodos mais secos e úmidos cria um cenário que exige uma gestão de recursos hídricos equilibrada e a necessidade de adaptação das atividades dependentes das condições climáticas à essa variabilidade.

Em Souza, o SPI-24 oscila de valores negativos, indicando seca, para valores positivos, sugerindo períodos mais úmidos, ao longo dos dois anos.

Figura 11 - Gráficos de linha das 6 estações referente às previsões do SPI 24

Fonte: Elaborado pelo autor, 2023.

4 CONCLUSÃO

Em síntese, a análise abrangente das precipitações pluviométricas e o subsequente estudo do Índice Padronizado de Precipitação (SPI) oferecem insights essenciais sobre os padrões climáticos de uma região, revelando uma interligação complexa entre fatores naturais e sociais. Através dos dados coletados e das observações realizadas, emerge uma compreensão profunda da variabilidade climática e de suas implicações, que transcendem meras estatísticas para impactar diretamente a vida das comunidades.

A análise das médias, desvios padrão e valores extremos das chuvas em seis diferentes estações pluviométricas proporcionou um panorama geográfico diversificado das tendências climáticas, destacando João Pessoa com a maior média de chuvas e Picuí com a menor. A visualização dos histogramas evidenciou que, apesar da prevalência de chuvas com valores menores, eventos excepcionais de alta intensidade não são raros, ressaltando a necessidade de planejamento resiliente para lidar com variações extremas.

Ao adentrar na análise das séries temporais, foi possível identificar anos específicos com precipitações notáveis em cada localidade. O SPI revelou sua eficácia em detectar eventos de chuvas e secas, proporcionando uma compreensão mais profunda da dinâmica climática em diferentes escalas temporais. Os anos marcados por chuvas intensas e secas severas variaram entre as estações, sublinhando a complexidade do clima regional e a importância de abordagens adaptativas para mitigar impactos adversos.

Através das diferentes escalas do SPI, ficou claro que eventos climáticos intensos podem afetar amplamente uma região, independentemente da escala temporal analisada. A década de 1990 e os anos 2015, 2016 e 2017 emergiram como um período de secas acentuadas, lançando luz sobre a necessidade de compreender e enfrentar tendências climáticas históricas para garantir uma preparação eficaz para eventos futuros.

A análise das previsões climáticas, em particular do SPI, para diversos municípios demonstra a profundidade e a complexidade do clima regional e suas implicações na gestão hídrica, agricultura e planejamento urbano. O modelo Extra Trees Regressor (ET) destacou-se em sua capacidade de interpretar essa variabilidade, embora persistam desafios na precisão e aplicabilidade das previsões. Esses resultados enfatizam a necessidade de uma abordagem holística na modelagem climática, combinando expertise técnica, entendimento climático e considerações sociopolíticas.

REFERÊNCIAS

ALFIAN, G.; SYAFRUDIN, M.; FAHRURROZI, I.; FITRIYANI, N. L.; ATMAJI, F. T. D.; WIDODO, T.; BAHYAH, N.; BENES, F.; RHEE, J. **Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method**. *Computers*, v. 11, p. 136, 2022. Disponível em: <<https://doi.org/10.3390/computers11090136>>. Acesso em: 28 de Setembro de 2023.

ALI, Moez. **PyCaret: An open source, low-code machine learning library in Python**. Abril 2020. Manual. Disponível em: <https://www.pycaret.org>. Nota: PyCaret version 1.0.

Breiman, L. **Random Forests**. *Machine Learning* 45, 5–32 (2001).
<https://doi.org/10.1023/A:1010933404324>

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random Forests. In: ZHANG, C.; MA, Y. (Ed.). **Ensemble Machine Learning: Methods and Applications**. [S.l.]: Springer, 2011. cap. 5, p. 157-176. DOI: 10.1007/978-1-4419-9326-7_5.

Da Silva, A.S.A., Menezes, R.S.C. and Stosic, T., 2021. **Análise multifractal do índice de precipitação padronizado**. *Research, Society and Development*, 10(7), pp.e24710716535-e24710716535.

De Farias¹, A.A., Souza, J.T.A. and de Sousa, F.D.A.S., 2014. **IDENTIFICAÇÃO E ANÁLISE DE SECAS SEVERAS E EXTREMAS NO MUNICÍPIO DE TAPEROÁ-PB**. *Revista Brasileira de Geografia Física*, vol.07, n.05 (Número Especial- VIWMCRHPE), 818-826

De Souza Guedes, R.V., Macedo, M.J.H. and de Sousa, F.D.A.S., 2012. **Análise espacial de eventos de secas com base no índice padronizado de precipitação e análise de agrupamento**. *Brazilian Journal of Environmental Sciences (Online)*, (23), pp.55-65.

DIDÁTICA TECH. **Como funciona o algoritmo ExtraTrees**. Disponível em: <<https://didatica.tech/como-funciona-o-algoritmo-extratrees/>>. Acesso em: 28 de Setembro de 2023.

FILHO, M. **MAE (Erro Médio Absoluto) Em Machine Learning**. Disponível em: <<https://mariofilho.com/mae-erro-medio-absoluto-em-machine-learning/>>. Acesso em: 28 de Setembro de 2023.

Friedman, Jerome H. **Greedy Function Approximation: A Gradient Boosting Machine.**” *The Annals of Statistics* 29, no. 5 (2001): 1189–1232. <http://www.jstor.org/stable/2699986>.

Geurts, P., Ernst, D. & Wehenkel, L. **Extremely randomized trees**. *Mach Learn* 63, 3–42 (2006). <https://doi.org/10.1007/s10994-006-6226-1>

HARUN. **Metrics - MSE, R², RMSLE**. 2019. Disponível em: <https://hrngok.github.io/posts/metrics/>. Acesso em: 06 out. 2023.

Hasan, H., Radi, N.A. and Kassim, S., 2012, July. **Modeling of extreme temperature using generalized extreme value (GEV) distribution: A case study of Penang**. In *Proceedings of the world congress on engineering* (Vol. 1, pp. 181-186).

ICMC JÚNIOR. **Random Forest**. Disponível em: <https://icmcjunior.com.br/random-forest/>. Acesso em: 28 de Setembro de 2023.

Macedo, M.J.H., de Souza Guedes, R.V. and de Assis Salviano, F.D.A., 2011. **Monitoramento e intensidade das secas e chuvas na cidade de Campina Grande/PB**. *Revista Brasileira de Climatologia*, 8.

Macedo, M.J.H., de Souza Guedes, R.V., de Sousa, F.D.A.S. and da Cunha Dantas, F.R., 2010. **Análise do índice padronizado de precipitação para o estado da Paraíba, Brasil**. *Ambiente & Água-An Interdisciplinary Journal of Applied Science*, 5(1), pp.204-214.

MASUI, T. **All You Need to Know about Gradient Boosting Algorithm – Part 1**. Regression. Towards Data Science, [s.l.], 20 jan. 2022. Disponível em: <https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-1-regression-2520a34a502>. Acesso em: 28 de Setembro de 2023.

Mathbout, S., Lopez-Bustins, J.A., Martin-Vide, J., Bech, J. and Rodrigo, F.S., 2018. **Spatial and temporal analysis of drought variability at several time scales in Syria during 1961–2012**. *Atmospheric Research*, 200, pp.153-168.

MATPLOTLIB. **Documentação estável**. Disponível em: <https://matplotlib.org/stable/index.html>. Acesso em: 20 ago. 2023.

McKee, T.B., Doesken, N.J. and Kleist, J., 1993, January. **The relationship of drought frequency and duration to time scales**. In *Proceedings of the 8th Conference on Applied Climatology* (Vol. 17, No. 22, pp. 179-183).

OTCHERE, D. A.; GANAT, T. O. A.; OJERO, J. O.; TACKIE-OTOO, B. N.; TAKI, M. Y. **Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions**. *Journal of Petroleum Science and Engineering*, v. 208, Part E, p. 109244, 2022. ISSN 0920-4105. Disponível em: <https://doi.org/10.1016/j.petrol.2021.109244>. Acesso em: 28 de Setembro de 2023.

PADHMA, M. **End-to-End Introduction to Evaluating Regression Models**. Analytics Vidhya, 2023. Disponível em: <https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models/>. Acesso em: 06 out. 2023.

PANDAS. **Documentação**. Disponível em: <https://pandas.pydata.org/docs/>. Acesso em: 20 ago. 2023

PARAÍBA TOTAL. **Geografia**. Disponível em: <https://www.paraibatotal.com.br/a-paraiba/geografia/>. Publicado em: 19 jul. 2021. Acesso em: 20 ago. 2023.

PYTHON SOFTWARE FOUNDATION. **Documentação Python 3.x** (Português do Brasil). Disponível em: <https://docs.python.org/pt-br/3/>. Acesso em: 20 ago. 2023.

Santos, C.A.G., Brasil Neto, R.M., da Silva, R.M. and dos Santos, D.C., 2019. **Innovative approach for geospatial drought severity classification: a case study of Paraíba state, Brazil**. *Stochastic Environmental Research and Risk Assessment*, 33(2), pp.545-562.

SEABORN. **Introdução ao tutorial**. Disponível em: <https://seaborn.pydata.org/tutorial/introduction.html>. Acesso em: 20 ago. 2023.

Sodré, G. d. A. 2040: **o aquecimento global gerando impacto econômico na produção de cana no Estado de São Paulo**. Tese (Doutorado), 2010. Svoboda, M., Hayes, M. and Wood, D., 2012. Standardized precipitation index: user guide.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**. Addison Wesley, 2005. ISBN: 0321321367.

TEAM, R. C. R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2013. [S.l.]: ISBN 3-900051-07-0, 2014.

Thom, H.C., 1958. **A note on the gamma distribution**. *Monthly weather review*, 86(4), pp.117-122.

Wang, K.Y., Li, Q.F., Yang, Y., Zeng, M., Li, P.C. and Zhang, J.X., 2015. **Analysis of spatio-temporal evolution of droughts in Luanhe River Basin using different drought indices**. *Water Science and Engineering*, 8(4), pp.282-290.

XARRAY. **Documentação estável**. Disponível em: <https://docs.xarray.dev/en/stable/>. Acesso em: 20 ago. 2023.