



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS SOBRAL
CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

KAMILA AMÉLIA SOUSA GOMES

**ANÁLISE DOS DADOS OBTIDOS PELOS SENSORES DE TRÂNSITO ATRAVÉS DE
SÉRIES TEMPORAIS**

SOBRAL

2021

KAMILA AMÉLIA SOUSA GOMES

ANÁLISE DOS DADOS OBTIDOS PELOS SENSORES DE TRÂNSITO ATRAVÉS DE
SÉRIES TEMPORAIS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do *Campus* Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Iális Cavalcante de Paula Júnior

SOBRAL

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

G614a Gomes, Kamila Amélia Sousa.
Análise dos dados obtidos pelos sensores de trânsito através de séries temporais / Kamila Amélia
Sousa Gomes. – 2021.
50 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Sobral,
Curso de Engenharia da Computação, Sobral, 2021.
Orientação: Prof. Dr. Iális Cavalcante de Paula Júnior.

1. Séries Temporais. 2. Sensores de Trânsito. 3. Anomalias. 4. Prophet. I. Título.

CDD 621.39

KAMILA AMÉLIA SOUSA GOMES

ANÁLISE DOS DADOS OBTIDOS PELOS SENSORES DE TRÂNSITO ATRAVÉS DE
SÉRIES TEMPORAIS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do *Campus* Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Iális Cavalcante de Paula
Júnior (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Régis Pires Magalhães
Universidade Federal do Ceará(UFC)

Prof. Me. David Nascimento Coelho
Universidade Federal do Ceará(UFC)

Aos meus avós: Maria (*in memoriam*), Joaquim (*in memoriam*), Messias (*in memoriam*) e Raimunda. E aos meus pais, os grandes incentivadores dos meus sonhos.

AGRADECIMENTOS

Esta seção, certamente, é uma das mais importantes, pois me permite vivenciar um dos maiores sentimentos do mundo: gratidão!

Grata à Deus, pela força, amparo, sabedoria e graças concedidas durante o decorrer de minha graduação. E, principalmente, por ter sido fonte de esperança nos momentos difíceis.

Aos meus pais, Luzinete e Epaminondas, pelo apoio incondicional, inspiração e encorajamento. Mãe, esse sonho também é seu! E aos meu irmãos, Jorge Rafael e João Guilherme, por me motivarem a ser melhor todos os dias. Obrigada por acreditarem em mim, quando até mesmo eu, duvidava.

Aos professores do curso de Engenharia da Computação e aos colaboradores da instituição UFC, pelo comprometimento. Ao prof. Dr. Wendley, pela assistência e ensinamentos me proporcionado. À Michelle, pela serenidade e disponibilidade de sempre. Em especial, ao meu orientador Iális, pelo exemplo de ser humano, incentivo, paciência e dedicação ao curso.

Aos professores participantes da banca examinadora, por aceitarem o convite, dedicando tempo e atenção ao trabalho.

Ao apoio financeiro da universidade que me permitiu a conclusão do curso.

Aos amigos e colegas que tornaram essa caminhada menos árdua. Aos companheiros de vida Ana, Sandra, Eduardo e Jean, obrigada fazerem parte da minha vida, vocês são fundamentais. À minha “panelinha”: Hugo, Ângela e Lucas Costa, por compartilharem os sofrimentos do cotidiano da faculdade, pelas noites de estudos e pizzas de consolação. Ao Expedito, por dividir comigo alegrias e aflições. Aos demais que não foram citados, mas sabem o valor e a importância que possuem, vocês foram indispensáveis durante essa realização.

Por fim, à todos que participaram, de forma direta ou indireta, da concretização deste sonho. Muito obrigada!

“O tamanho dos seus sonhos deve sempre exceder a sua capacidade de alcançá-los. Se os seus sonhos não te assustam, eles não são grandes o suficiente.”

(Ellen Johnson-Sirleaf)

RESUMO

A evolução das tecnologias digitais permitem que as técnicas e ferramentas de monitoramento do trânsito de veículos tornem-se mais eficientes. Isto porquê, seria inviável um operador humano lidar sozinho com tantas informações. No entanto, podem ocorrer problemas durante a captura dessas informações e danificá-las. À essas informações errôneas ou insuficientes, que não seguem um comportamento padrão se comparada ao comportamento da maioria dos dados, dá-se o nome de anomalias. E, para a detecção desses objetos não esperados, pode-se utilizar séries temporais. São chamadas de séries temporais um conjunto de observações feitas sequencialmente ao longo do tempo. Sabendo disso, o objetivo do trabalho é analisar os dados capturados pelos sensores de trânsito através de séries temporais. Em seu desenvolvimento serão abordados conceitos teóricos fundamentais para o entendimento da metodologia, identificação de padrões anômalos na base de dados e aplicação de métodos de predição como: *Arima*, *Prophet*, *Simple Exponential Smoothing*, *Walk Forward* e *Random Walk*. Por fim, comparou-se os métodos através das métricas de erro: *Root Mean Squared Error* (RMSE) e *Mean Absolute Error* (MAE). Em síntese, foi observado que, o método de predição *Prophet* apresentou melhor desempenho em relação as demais técnicas.

Palavras-chave: Séries Temporais. Sensores de Trânsito. Anomalias. Prophet

ABSTRACT

The evolution of digital technologies allows techniques and tools for monitoring vehicle traffic to become more efficient. This is why, it would not be viable for a human operator to deal with so much information alone. However, problems can occur while capturing information and damaging it. This erroneous or insufficient information, which does not follow a standard behavior when compared to the behavior of most data, is called anomalies. And, for the detection of these unexpected objects, time series can be used. A series of observations made sequentially over time are called time series. Knowing this, the objective of the work is to analyze the data captured by the traffic sensors through time series. In its development, fundamental theoretical concepts for understanding the methodology, identification of anomalous patterns in the database and application of prediction methods such as: *Arima*, *Prophet*, *Simple Exponential Smoothing*, *Walk Forward* and *Random Walk*. Finally, the methods were compared using error metrics: RMSE and MAE. In summary, it was observed that the *Prophet* prediction method performed better than the other techniques.

Keywords: Time Series. Traffic Sensors. Anomalies. Prophet.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ilustração de dados anômalos em dois formatos de classificação. À esquerda, tem-se a classificação por <i>clusters</i> . À direita, tem-se a classificação por série temporal.	19
Figura 2 – Representação de Anomalias Pontuais.	20
Figura 3 – Representação de uma Anomalia Contextual.	21
Figura 4 – Representação de uma Anomalia Coletiva.	21
Figura 5 – Exemplificação do Resultado do Teste <i>Augmented Dickey Fuller</i> (ADCF) . .	24
Figura 6 – Terminologia ARIMA.	26
Figura 7 – Funções ACF e PACF.	26
Figura 8 – Exemplificação <i>Walk Forward</i>	28
Figura 9 – Fluxograma da Metodologia.	33
Figura 10 – Não-Estacionaridade da Série	35
Figura 11 – Aplicação do Teste ADCF na Série.	35
Figura 12 – Aplicação do Teste ADCF após Transformação	36
Figura 13 – Aplicação das Funções ACF e PACF na série.	36
Figura 14 – Distribuição de frequência das velocidades registradas.	38
Figura 15 – Comparação dos registros antes e depois da limpeza de dados.	40
Figura 16 – SES com parâmetro de suavização $\alpha = 0.2$	41
Figura 17 – SES com parâmetro de suavização $\alpha = 0.7$	41
Figura 18 – SES com parâmetro de suavização α automatizado.	42
Figura 19 – Resultado Saída <i>Prophet</i> com dados temporais diários.	42
Figura 20 – Gráfico <i>Prophet</i> com dados temporais diários.	43
Figura 21 – Resultado Saída <i>Prophet</i> com dados temporais por hora diariamente.	43
Figura 22 – Gráfico <i>Prophet</i> com dados temporais por hora diariamente.	44
Figura 23 – Gráfico <i>Random Walk 1D</i>	44
Figura 24 – Gráfico <i>Walk Forward</i>	45
Figura 25 – Comparação resultado RMSE.	45
Figura 26 – Comparação resultado MAE.	46

LISTA DE TABELAS

Tabela 1 – Sensores Potencialmente com Falhas	39
Tabela 2 – Comparação de Resultados do Método Arima	40
Tabela 3 – Comparação de Resultados dos Métodos de Predição e Métricas de Erro . .	46

LISTA DE ABREVIATURAS E SIGLAS

ABEETRANS	Associação Brasileira das Empresas de Engenharia de Trânsito
ACF	Função de Autocorrelação
ADCF	<i>Augmented Dickey Fuller</i>
ARIMA	Modelo Auto Regressivo Integrado de Médias Móveis
CGU	Controladoria Geral da União
Funceme	Fundação Cearense de Meteorologia e Recursos Hídricos
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
PACF	Função de Autocorrelação Parcial
RMSE	<i>Root Mean Squared Error</i>
SES	<i>Simple Exponential Smoothing</i>
UFC	Universidade Federal do Ceará

LISTA DE SÍMBOLOS

α	Parâmetro de Suavização
β	Parâmetro de Crescimento
γ	Parâmetro de Sazonalidade
\in	Pertence
Δ	Delta

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Motivação	16
1.2	Objetivos	17
1.2.1	<i>Objetivo Geral</i>	17
1.2.2	<i>Objetivos Específicos</i>	17
1.3	Organização do Trabalho	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Anomalias	19
2.1.1	<i>Tipos de Anomalias</i>	20
2.2	Séries Temporais	22
2.2.1	<i>Estacionaridade</i>	23
2.3	Bibliotecas	24
2.4	Limpeza de Dados	25
2.5	Métodos de Predição	25
2.5.1	<i>ARIMA</i>	25
2.5.2	<i>Prophet</i>	27
2.5.3	<i>Walk Forward</i>	28
2.5.4	<i>Simple Exponential Smoothing</i>	29
2.5.5	<i>Random Walk</i>	29
2.6	Métricas de Validação de Erros	30
2.6.1	<i>Mean Absolute Error</i>	30
2.6.2	<i>Root Mean Squared Error</i>	30
3	TRABALHOS RELACIONADOS	32
4	METODOLOGIA	33
4.1	Dados dos Sensores	33
4.2	Pré-Processamento de Dados	33
4.3	Limpeza de Dados	34
4.4	Métodos de Predição	35
5	RESULTADOS	38
5.1	Análise de Sensores	38

5.2	Métodos de Predição	40
5.3	Análise de Resultados	45
6	CONCLUSÃO E TRABALHOS FUTUROS	47
	REFERÊNCIAS	48

1 INTRODUÇÃO

Com o aumento de veículos automotores, o fluxo de trânsito torna-se cada vez mais saturado. Logo, com a crescente ocorrência de acidentes ligados a alta velocidade e a dificuldade dos agentes de trânsito de registrá-las visualmente, houve a necessidade de desenvolver mecanismos de suporte à segurança de forma automática e confiável. Assim, surgiram os radares como os primeiros equipamentos de fiscalização eletrônica (MING, 2006).

O avanço da tecnologia permite uma maior eficiência para orientar e fiscalizar o tráfego. Hoje existem: sinalização dinâmica, comunicação com aplicativos de navegação, uso de *drones* na fiscalização rodoviária, sistemas inteligentes que averiguam grandes volumes de dados em busca de padrões que direcionem melhor as intervenções, além de radares e câmeras mais eficientes. Ademais, as câmeras são essenciais em fiscalizar o fluxo de veículos, indicar congestionamentos, reconhecer veículos irregulares e condutas suspeitas (REVISTA SIM, 2018).

No entanto, por localizarem-se em pontos fixos, determinadas vezes, os sensores não capturam imagens nítidas e impedem o registro sobre a passagem do veículo. A Controladoria Geral da União (CGU) através de uma pesquisa, concluiu que algumas características sobre placas ilegíveis se referem à fotos escuras, reflexo do sol, ou chuvas. Além disso, a captura de dados é fundamental para geração de informações que possibilitem a melhora na segurança no trânsito (SIQUEIRA, 2015).

Outro grande aliado para o trabalho dos agentes de segurança pública é o *Big Data*. Segundo Newman Marques da Silva, diretor de tecnologia da ABEETRANS, na Europa já existe uma grande infraestrutura de coleta de dados no trânsito. Esses dados chegam aos sistemas em um fluxo volumoso de informações brutas, atualizadas a cada segundo. Para que esses dados sejam produtivos no gerenciamento de trânsito, é preciso filtrar elementos relevantes e buscar padrões de comportamento para identificar diversas situações nas rodovias (ABEETRANS, 2016).

Nesse contexto, uma das contrariedades que dificultam a precisão da fiscalização são as anomalias. De acordo com Nova Filho e Pessoa (2015) anomalias podem ser vistas como *outliers*. E, assim como o *outlier* desvia-se de forma relevante dos objetos restantes, por não apresentar características similares ou por ter sofrido interferência externa ao processo normal de seu domínio, assim também são as anomalias. Outra definição para este conceito são padrões que não apresentam concordância com o comportamento normal. Dessa forma, a finalidade da detecção de anomalias é encontrar objetos não esperados (CARVALHO *et al.*, 2018).

Para Santos e Lins Júnior (2018) são múltiplos fatores responsáveis pelas anomalias, como interferência de outros equipamentos na mesma faixa de frequência, falha de hardware e atividades maliciosas. Nova Filho e Pessoa (2015) afirmam que provavelmente as anomalias sejam introduzidas nos dados por: erros de coleta, variação natural - caracterizada por variação inerente do dado - ou dados de classes distintas. No entanto, em diversos casos, seu surgimento ainda é desconhecido.

Para auxiliar na detecção de anomalias, no reconhecimento desses padrões ou em alguma mudança de comportamento podemos utilizar séries temporais (KUPLICH *et al.*, 2013). Uma série temporal é definida por Chino (2014) como "uma coleção de m observações realizadas sequencialmente ao longo do tempo". Dados que descrevem mudanças de uma variável ao longo do tempo desde que esse tempo tenha a mesma periodicidade, é uma série temporal. O tempo pode apresentar-se de forma: diária, mensal, semestral ou anual. Como exemplo comuns, podemos citar: o consumo diário de energia elétrica de uma região, vendas mensais de uma empresa, a cotação do preço de uma ação na bolsa de valores (CAMPOS *et al.*, 2006).

As séries temporais podem ser utilizadas em diversos contextos. Em CAMPOS *et al.* (2006) elabora-se um modelo de previsão para o preço do frango inteiro resfriado em São Paulo, utilizando a metodologia de previsões de séries temporais ARIMA. Em MARTIN *et al.* (2016) o trabalho teve como objetivo pesquisar o crescimento do número de automóveis no município de Joinville em Santa Catarina por meio da análise de séries temporais, onde foram propostos o modelo ARIMA e modelos de suavização exponencial. Dessa forma, estudos dessa natureza são fundamentais para verificar o padrão de comportamento dos dados ao longo do tempo, com intuito que sejam identificados possíveis problemas com antecedência.

1.1 Motivação

A tecnologia é um grande aliada da segurança pública. Hoje, grandes volumes de dados são gerados por câmaras, radares, entre outros equipamentos de rodovias. A grande vantagem disso é o fortalecimento da segurança pública com a redução do roubo de veículos, do crime e de acidentes e mortes no trânsito. Um exemplo na realidade do sistema de trânsito brasileiro é o "FotoSiga Segurança Pública", desenvolvido pela empresa Fotosensores de Fortaleza, associada a Associação Brasileira das Empresas de Engenharia de Trânsito (ABEETRANS). O FotoSiga trata-se de um sistema de monitoramento de veículos, mediante a detecção de sua passagem. Além disso, com base na coleta das informações, é possível detectar padrões de comportamento

desses veículos (ABEETRANS, 2016).

Através do padrão de conduta, nota-se comportamentos anormais ou indesejáveis, chamados de anomalias. É importante analisar as alterações no comportamento das variáveis para que seja possível corrigir problemas e prevenir erros (FIGUEIRÊDO *et al.*, 2020). Freitas *et al.* (2019) define objeto anômalo aquele que tem distanciamento de forma significativa do restante dos dados, se desviando do padrão. As razões desse acontecimento podem ser desconhecidas, mas Santos e Lins Júnior (2018) sugerem como causa alguma falha no hardware e Nova Filho e Pessoa (2015) mencionam erros na coleta de dados.

Neste trabalho serão aplicadas conhecimentos na área de séries temporais para analisar os padrões anômalos nos sensores de trânsitos. Séries temporais são definido por Ehlers (2007) como coleção de observações feitas em um conjuntos de dados ao percorrer do tempo. É importante ressaltar que, existe uma dependência dessas variáveis com experiencias anteriores e o foco é analisar e modelar esta dependência. As séries temporais fazem parte de uma área multidisciplinar e podem ser empregadas em diversos campos do conhecimento, como: Economia, Epidemiologia, Meteorologia, Medicina, entre tantos outros.

Dessa forma, a finalidade desta pesquisa é primordial para auxiliar nas possíveis falhas dos dados capturados pelos sensores de trânsito e fortalecimento na função dos agentes de segurança pública, com informações e ferramentas na solução e prevenção de indícios de atividade criminosa ocorrida, por exemplo.

1.2 Objetivos

Nesta seção, será abordada a ideia central desse trabalho, através da descrição de seus objetivos.

1.2.1 Objetivo Geral

Este trabalho tem como finalidade verificar a qualidade das informações capturadas nos sensores de trânsito através de séries temporais, reconhecendo padrões anômalos e implementando técnicas preditivas para a análise de seu comportamento.

1.2.2 Objetivos Específicos

- Explorar trabalhos relacionados à literatura de séries temporais em sensores de trânsitos;

- Identificar comportamentos anômalos nos sensores disponibilizados no conjunto de dados utilizado no trabalho;
- Avaliar a eficiência de diferentes técnicas de predição de séries temporais em sensores de tráfegos;
- Validar a metodologia proposta no projeto mediante os experimentos e comparar os resultados das técnicas de análise de séries temporais.

1.3 Organização do Trabalho

O desenvolvimento deste trabalho deu-se através de 6 capítulos. O Capítulo 2 apresenta uma revisão bibliográfica contendo, de forma resumida e geral, os conceitos teóricos necessários para ampliação da pesquisa. O capítulo 3 traz trabalhos relacionados a temática abordada. O capítulo 4, referente a metodologia, é dividido em duas partes: identificar sensores anômalos e, em seguida, aplicar modelos de predição. Entre esses modelos têm-se: *Arima*, *Prophet*, *Exponential Smoothing*, *Walk Forward* e *Random Walk*. A avaliação do desempenho dessas técnicas será apresentada no Capítulos 5, mostrando os resultados de previsões e comparando-os. Por fim, o capítulo 6 disponibiliza uma breve discussão do que foi abordado, conclusões e propostas futuras.

2 FUNDAMENTAÇÃO TEÓRICA

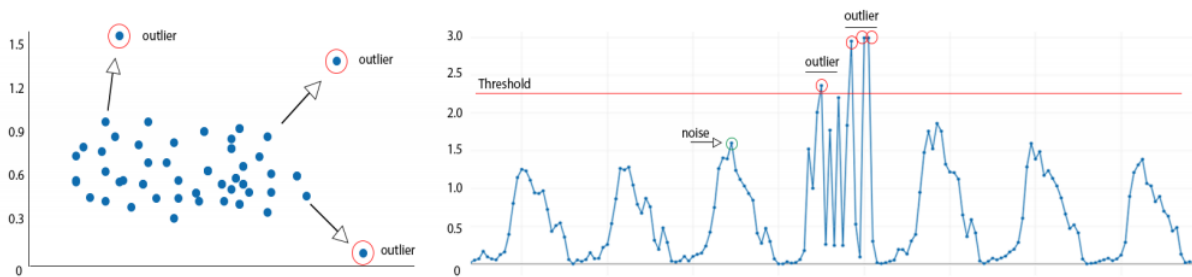
Neste capítulo é apresentado o levantamento bibliográfico relacionado ao desenvolvimento deste trabalho através das subseções a seguir.

2.1 Anomalias

Anomalia é definido por Freitas *et al.* (2019) como um processo estatístico aplicado a um conjunto de dados onde são observados pontos que se desviam do padrão. E, para Santos e Lins Júnior (2018) esses modelos não seguem um padrão por diversos motivos, como falha ou interferência nos hardwares ou equipamentos.

Em seu trabalho, Mata (2017) expõe a representação de duas figuras anômalas. Na Figura 1, tem-se à esquerda, a representação de pontos na forma de dados agrupados, ou *clusters*. Nela, os pontos mais afastados são denominados anômalos. E, à direita, têm-se a representação na forma de série temporal. Nesse caso, são anômalos aqueles que se encontram acima de um *threshold*(limite) específico.

Figura 1 – Ilustração de dados anômalos em dois formatos de classificação. À esquerda, tem-se a classificação por *clusters*. À direita, tem-se a classificação por série temporal.



Fonte – Mata (2017).

A Figura 1 traz a identificação de anomalias, que têm sido cada vez mais relevante para diversas áreas. No entanto, não existe uma abordagem global e a cada situação devem ser aplicadas diferentes técnicas. Além disso, definir o que é um comportamento padrão ou incomum não é tarefa trivial, isto porquê existem distribuições que podem induzir que determinados pontos sejam anômalos, mesmo que estes na realidade não sejam (MATA, 2017).

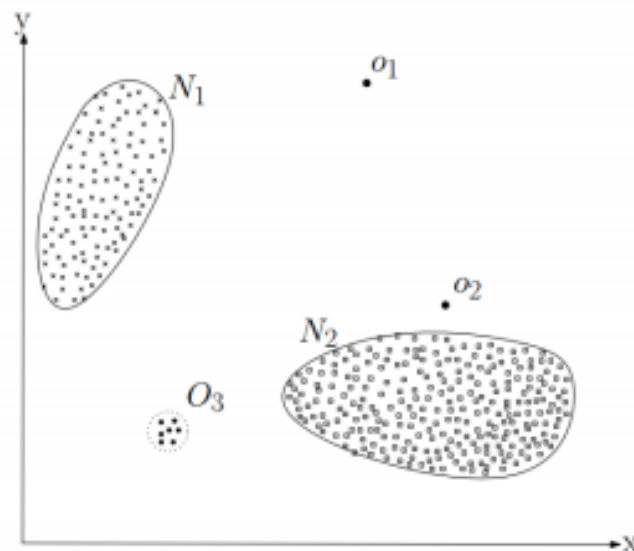
2.1.1 Tipos de Anomalias

"Determinar o tipo de anomalia sendo identificado é um aspecto importante das técnicas de detecção de anomalias"(GONZAGA, 2017). Classificam-nas em três tipos: Pontuais, Contextuais e Coletivas.

1. Anomalia Pontual:

A anomalia pontual, também conhecida como anomalia global, ocorre quando uma observação particular desvia-se consideravelmente das demais. Ela é o foco da maior parte das pesquisas e aplicações (FREITAS *et al.*, 2019). A Figura 2, representa uma anomalia pontual.

Figura 2 – Representação de Anomalias Pontuais.



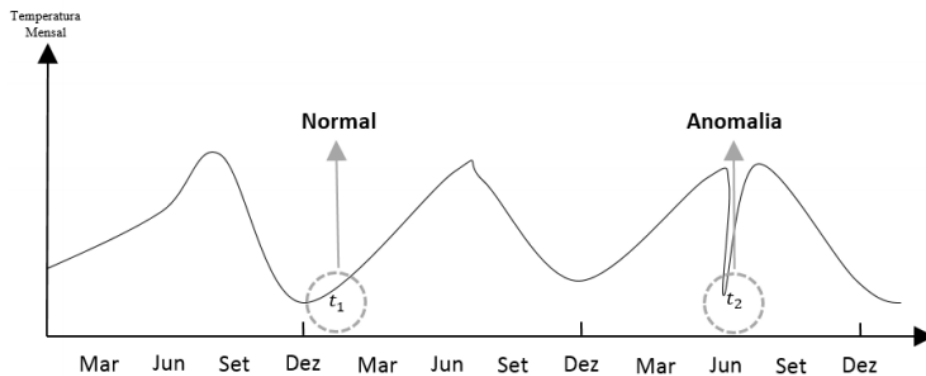
Fonte – Gonzaga (2017).

A Figura 2 apresenta três conjuntos de dados: N1, N2 e O3, e outros dados individuais: O1 e O2. Considera-se os pontos O1 e O2 exemplos de anomalias pontuais.

2. Anomalia Contextuais:

As Anomalias Contextuais são consideradas como *outliers* condicionais, pois somente é considerada anômala em um contexto específico. Nesse tipo, a instância é analisada por atributos comportamentais e textuais (FREITAS *et al.*, 2019). Em seu trabalho, Gonzaga (2017) traz um exemplo: suponha que um sensor registre altas temperatura no decorrer do dia e essa temperatura caiu durante a noite. Nesse caso, ocorreu um comportamento fora do contexto, logo uma anomalia contextual. Como apresentado na Figura 3.

Figura 3 – Representação de uma Anomalia Contextual.



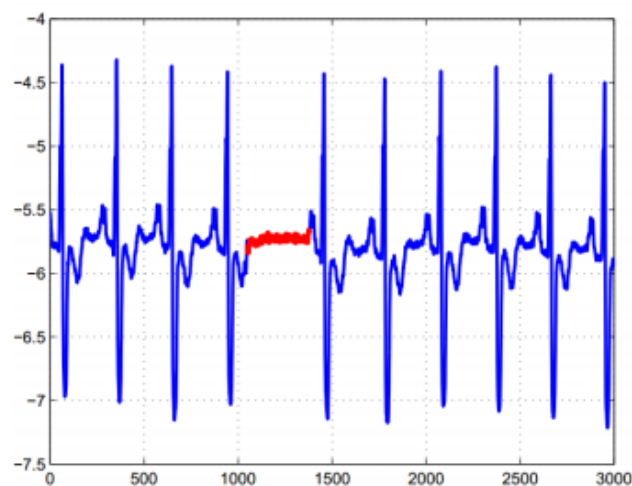
Fonte – Freitas *et al.* (2019)

A Figura 3 ilustra um caso de anomalia contextual. Embora as temperaturas t_1 e t_2 possam ter mesmo valor, devido ao seu contexto, t_2 é uma anomalia.

3. Anomalia Coletiva:

Por fim, a Anomalia Coletiva ocorre quando em um conjunto de dados, existe um subconjunto anômalo em relação ao restante das observações. Um caso em que podemos encontrar esse tipo de anomalia é no eletrocardiograma humano (GONZAGA, 2017). Como mostrado na Figura 4.

Figura 4 – Representação de uma Anomalia Coletiva.



Fonte – Gonzaga (2017)

Na Figura 4, nota-se que se observados individualmente, não são considerados anômalos os valores da região vermelha. No entanto, quando observado no geral, ocorrência em sequência indica anomalia nessa região.

2.2 Séries Temporais

"Uma Série Temporal é uma coleção de observações feitas sequencialmente ao longo do tempo"(EHLERS, 2007). Ou seja, são dados coletados em intervalos regulares de tempo de um período amostral sobre observações de determinada variável (NOVA FILHO; PESSOA, 2015).

A Série Temporal é dada por um conjunto de observações $Y(t)$, $t \in T$, sendo Y a variável de interesse e T o conjunto de índices (MIGON, 2007). Matematicamente, pode ser representada na forma discreta, como na Equação 2.1.

$$T = \{t_1, t_2, \dots, t_n\} \quad (2.1)$$

Pode-se citar, a título de exemplo para as séries temporais, séries ordenadas cronologicamente, como: vendas médias de determinado item, Produto Interno Bruto (PIB), valores de exportações, temperatura, a taxa de inflação, entre tantos outros (MARTIN *et al.*, 2016).

Souza (2017) afirma que as séries temporais podem ser caracterizadas pela combinação de componentes de: tendência, sazonalidade, cíclicos e erro.

1. Tendência: O comportamento de tendência é aquele de longo prazo da série, que são responsáveis por mudanças graduais de modo que se manifestam por meio de movimentos continuados, que podem ser ascendentes, descendentes ou estáveis. Por exemplo, é possível que ela seja ocasionada pela mudança gradual de hábitos de consumo, pelo crescimento demográfico, ou através de aspectos que alterem a variável de interesse no período de tempo (SOUZA, 2020).
2. Sazonalidade: Já a sazonalidade, pode ser identificada de maneira visual, pois faz referência a mudanças de comportamentos que ocorrem continuamente em um determinado período, no qual essas oscilações - subidas e quedas - se repetem em épocas específicas do ano. Citando um caso análogo, temos que o número de passageiros que utilizam o transporte aéreo geralmente é maior em períodos de férias do que nos demais meses do ano (OLIVEIRA, 2019a).
3. Cíclicos: Esse componente refere-se aos ciclos da série que, assim como os componentes sazonais, se repetem com certa periodicidade, no entanto se apresentam em prazos maiores e menos regulares, como fenômenos climáticos (SOUZA, 2020).

4. Erro: Por fim, o componente de erro é dado pelas alterações irregulares da série, que não são explicados pelas variações de tendência, sazonalidade ou cíclicas (SOUZA, 2017).

Em relação à variabilidade das observações, podem-se classificar as séries temporais em estacionárias e não estacionárias (OLIVEIRA, 2019a). Uma série é considerada estacionária quando progride em torno de uma média constante, se apresentando de forma estável. No entanto, a grande maioria se apresenta de forma não estacionária, que é quando eles manifestam sazonalidades e tendências. Nessa condição, é necessário o uso de técnicas adequadas a tal situação (BEZERRA, 2006). Outra forma de se entender uma série temporal, do ponto de vista econométrico, é que elas podem ser univariadas e multivariadas. Considera-se univariada uma série baseada em uma única série histórica. Por conseguinte, considera-se uma série multivariada aquelas que envolvem mais de uma série histórica (SOUZA, 1981).

2.2.1 Estacionaridade

Uma série é considerada estacionária se suas médias e variâncias se mantêm constantes com o passar do tempo e a função de auto covariância apenas necessita da defasagem entre os instantes de tempo, mantendo um "equilíbrio" (WERNER; RIBEIRO, 2003).

Para iniciar uma modelagem preditiva é importante analisar se as propriedades estatísticas são constantes, ou seja, se a média, a variância e a autocorrelação mantêm-se ao longo do tempo. Caso ela alterne de acordo com o tempo, a série é considerada não estacionária. Quando a variância é pequena em relação à média, a realização de uma predição é bem mais simples, pois a média não se afasta significativamente. Além disso, é necessário verificar a correlação das variáveis, que acontece quando duas variáveis tem variação semelhante em relação ao desvio padrão (RABELO, 2019).

Verifica-se a estacionariedade da série através de alguns métodos, como: observar os gráficos (verificando visualmente se existe uma sazonalidade ou tendência óbvia), por meio de estatísticas resumidas (verificando se há diferenças significativas) ou por testes estatísticos. O mais utilizado é o teste estatístico de *Augmented Dickey Fuller* (BROWNLEE, 2016). Brownlee (2016) explica esse teste por meio de uma hipótese nula ou alternativa. A hipótese nula (H_0) do teste é quando a série temporal pode ser representada por uma raiz unitária, que não é estacionária. Por consequência, a hipótese alternativa (H_1) é que a série temporal é estacionária. Esse resultado é interpretado usando o valor p do teste. Um valor de p acima de um limite (como 5% ou 1%) admite hipótese nula, caso contrário, rejeita-se a hipótese nula. A seguir, na Figura 5, está um

exemplo do resultado de cálculo do teste *Augmented Dickey Fuller*:

Figura 5 – Exemplificação do Resultado do Teste ADF

Results of Dickey-Fuller Test:	
Test Statistic	-3.684896
p-value	0.004331
#Lags Used	26.000000
Number of Observations Used	1885.000000
Critical Value (10%)	-2.567587
Critical Value (5%)	-2.863075
Critical Value (1%)	-3.433824

Fonte – Arzamendia (2019)

Através da Figura 5, é possível notar, através do resultado emitido pelo teste ADF, que a série é estacionária. Isso acontece porquê, o foco principal neste resultado são os números “Estatística de Teste” e “Valores Críticos”. Para que uma série seja estacionária, é preciso que a estatística de teste seja menor que todos os valores críticos. Assim o primeiro valor precisa ser menor do que o último. E quanto menor o valor da estatística de teste, melhor (ARZAMENDIA, 2019).

Caso a série temporal não seja estacionária, Arzamendia (2019) afirma que existem técnicas que podem ser aplicadas para estacionar uma série. As principais técnicas são a transformação e a diferenciação. A transformação é uma técnica de correção que ressalta mais valores altos do que valores menores. Ela reduz essa tendência por meio de raiz quadrada, raiz cúbica, logaritmo. A diferenciação lida com a tendência e sazonalidade, e funciona através do cálculo da diferença da observação em um determinado instante com a do ponto anterior.

2.3 Bibliotecas

Bibliotecas são soluções modernas que reduzem os códigos através de um conjunto de módulos e funções. Na linguagem *Python* existem mais de 137.000 bibliotecas disponíveis para auxiliar e beneficiar os desenvolvedores (OLIVEIRA, 2019b). Há inúmeras bibliotecas que ajudam os cientistas de dados a executar seu trabalho, no entanto as mais conhecidas e aplicadas neste trabalho são elas: *Pandas*, *Numpy*, *Matplotlib*, *Plotly*, *Seaborn* e *Statsmodels* (COUTINHO, 2020). Oliveira (2019b) afirma que a biblioteca *Numpy* oferece suporte para diferentes objetos de matriz multidimensional, onde tem como vantagem ser eficiente, rápida e boa gerenciadora de matrizes e *arrays*. O autor também cita a biblioteca *Matplotlib*, usada para aplicações matemáticas ou científicas que necessitam plotagem de gráficos bidimensionais. Já Coutinho (2020) define a biblioteca *Pandas* uma das mais importantes para a ciência de dados,

pois ela pesquisa, organiza, representa e manipula informações com simplicidade. A biblioteca *Seaborn* e *Plotly* também são utilizadas para visualização de dados, no entanto a segunda trabalha com visualizações interativas e refinadas. Zhang (2018) define a biblioteca *Statsmodels* como um módulo *Python* que oferece funções e classes para implementar diferentes modelos estatísticos.

2.4 Limpeza de Dados

Uma base de dados é formada por muitos dados que podem conter informações faltantes ou irrelevantes, por isso é fundamental a etapa de limpeza de dados. Esta etapa é responsável pela remoção de valores, preenchimento de dados ausentes, redução de ruídos, entre outras tarefas. Como uma causa desse problema, pode-se citar, a falha na coleta desses dados. E, como solução, tem-se a remoção de atributos nulos, o preenchimento do atributo faltante com a média dos valores ou preenchimento com o valor mais repetido no *dataset*. É indispensável ressaltar que existem diferentes tipos de estratégias para lidar com determinado tipo de dados. Não há técnica universal para a solução de todos os problemas (GOMES, 2019).

2.5 Métodos de Predição

De acordo com Becker (2010) os métodos de previsão de séries temporais são métodos quantitativos e baseiam-se em observações anteriores através do registro de informações de comportamento padrão da série. Comumente, os métodos de previsão fundamentam-se na suposição que observações anteriores se comportam de acordo com o padrão de comportamento da série temporal. Além disso, a grande vantagem da utilização de técnicas estatísticas para essas previsões, é o reconhecimento precoce de comportamento das variáveis (REIS, 2015). Dessa maneira, a seguir são listadas algumas técnicas de previsão aplicadas neste projeto.

2.5.1 ARIMA

A metodologia de Box-Jenkins ou Modelo Auto Regressivo Integrado de Médias Móveis (ARIMA) (do inglês, *Regressive Integrated Moving Averages*) é uma das técnicas mais abordadas no estudo de séries temporais (CAMPOS *et al.*, 2006). Bandura *et al.* (2019) afirma que o ARIMA trata-se de um modelo estatístico linear para análise de séries temporais. Sobretudo, o ARIMA usa autocorrelação e média móvel para efetuar uma predição (VASCONCELLOS, 2018). Para aplicação desse modelo, a série precisa ser estacionária, isso quer dizer que é necessário

que haja uma estabilidade dos dados ao longo do tempo, contendo uma distribuição constante em torno de uma média móvel (BANDURA *et al.*, 2019).

Modelos preditivos possuem características singulares e manifestam isso através de seus próprios parâmetros. Dessa forma, o ARIMA também possui seus próprios parâmetros, no qual chamamos de termos. A sigla ARIMA é representada pelas letras p, d e q. O “p” está para o “AR”, à medida que “d” está para “I” e “q” está para “MA” (VASCONCELLOS, 2018). Como presente na Figura 6.

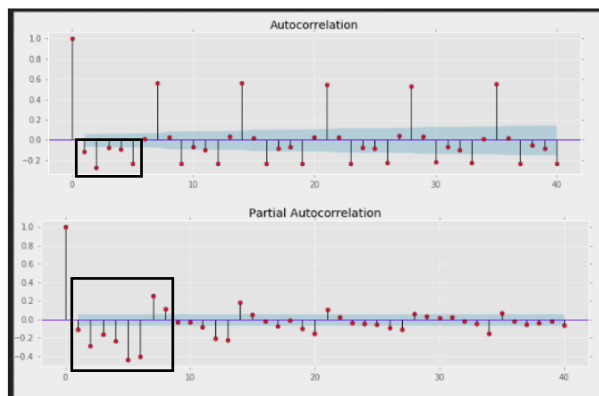
Figura 6 – Terminologia ARIMA.



Fonte – Adaptação de Vasconcellos (2018)

Na Figura 6, a letra “p” faz referência ao número de termos da parte auto regressiva. Isso possibilita agregar informações passadas ao modelo atual. Por exemplo, se nos três dias anteriores o tempo estava quente, é provável que amanhã também esteja. A letra “d” refere-se a integração ou número de diferenças não sazonais. Por fim, a letra “q” remete ao número de termos da média móvel. Permitindo estabelecer o erro do modelo como uma combinação linear dos valores de erro observados em momentos antecedentes (SOUZA, 2020). Para compreender o número ideal desses parâmetros para serem aplicados em seu modelo, é necessário entender Função de Autocorrelação (ACF) e a Função de Autocorrelação Parcial (PACF). Elas são exibidas por meio de um gráfico de pirulito (do inglês, *lollipop chart*). Como mostrado na Figura 7.

Figura 7 – Funções ACF e PACF.



Fonte – Vasconcellos (2018)

Através da Figura 7, para saber quantidade de termos AR utilizar, deve-se observar o gráfico de ACF e verificar se o segundo *lollipop* caiu exponencialmente. Se sim, devem ser contados o número de pirulitos que passam o valor crítico (faixa azulada horizontal) no gráfico de PACF. Acima, na Figura 7, o p teria oito termos. Em paralelo a isso, para a quantidade de termos q, verifique se o segundo lollipop do gráfico de PACF caiu exponencialmente. Se sim, devem ser contados o número de pirulitos que passam o valor crítico (faixa azulada horizontal) no gráfico de ACF. No caso da Figura 7, o q teria cinco termos (VASCONCELLOS, 2018). Após encontrar o número de termos ideal, o modelo já está pronto para ser treinado.

2.5.2 Prophet

O *Prophet* é um software de código aberto lançado pela equipe *Core Data Science do Facebook* e está disponível nas linguagens de programação: *Python* e *R* (TAYLOR; LETHAM, 2018). É uma biblioteca do *Facebook* usada para fazer previsão de dados de séries temporais baseada em um modelo aditivo, onde as tendências não lineares são adaptadas com sua sazonalidade, seja ela diária, semanal, anual ou, até mesmo, nos efeitos de feriados (PROPHET, 2017).

Lyla (2019) destaca algumas características destaques no *Prophet*:

- Preciso e rápido, usando um baixo poder computacional;
- Robusto para dados ausentes e alterações na tendência, e geralmente lida bem com *outliers*;
- Procedimento simples para ajuste (ajustes esses realizados através de parâmetros intuitivos que são fáceis de sintonizar);
- Inteiramente automatizado, ou seja, desempenha uma boa previsão sem muito esforço manual.

Taylor e Letham (2018) afirmam que, em sua essência, esse procedimento é um modelo de regressão aditivo decomposto com três principais componentes do modelo: tendência, sazonalidade e feriados. Como destacado na equação 2.2.

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (2.2)$$

Na Equação 2.2, a função $g(t)$ é uma curva de crescimento linear ou logística por partes para modelar mudanças não periódicas em séries temporais. A função $s(t)$ se refere a sazonalidade, através de mudanças periódicas. A função $h(t)$ representa os efeitos dos feriados, que acontecem em horários irregulares durante um ou mais dias. Por fim, o termo de *erro* ε

caracteriza quaisquer mudanças incomuns que não são condizentes no modelo (LYLA, 2019).

2.5.3 *Walk Forward*

Walk Forward, também conhecida como Janela Deslizante, é uma técnica de previsão aplicada quando trabalha-se com séries temporais usando médias móveis (SANTANA, 2020). O algoritmo *Walk Forward* tem início na parte mais à esquerda da série, e a cada iteração é calculado a média para gerar o resultado seguinte. Essa média é calculada de acordo com o tamanho da janela escolhida (CHINO, 2014). Como exemplo, utiliza-se a Figura 8.

Figura 8 – Exemplificação *Walk Forward*.



Fonte – Santana (2020)

Na Figura 8, os quadrantes representam uma porção dos dados que será utilizada para cálculo da média para geração do próximo dado. Dessa forma é percorrida toda a série seguindo uma ordem cronológica. Por exemplo, se em uma base de dados que contém a quantidade de vendas por mês, a cada três meses calcula-se a média móvel e usa-se o valor encontrado para prever o valor do mês seguinte (SANTANA, 2020).

O maior desafio quando trata-se de janela deslizante é como regular o tamanho da janela. Quando se trata de uma janela menor, o foco conserva-se no contexto atual, o que pode não representar dados para um contexto mais duradouro. Em contrapartida, uma janela maior poderá mesclar diferentes contextos. Dessa forma, podem ser ajustados e testados diferentes tamanhos para essa janela, de acordo com a aplicação. Alguns autores ainda estudam como utilizar esse tamanho de forma dinâmica (ALVES; ROSSI, 2018).

2.5.4 *Simple Exponential Smoothing*

A Suavização Exponencial, do inglês *Exponential Smoothing*, surgiu no final dos anos de 1950 e seus métodos de previsão são por meio de médias ponderadas exponencialmente de observações anteriores, com aplicação de pesos. Ou seja, o maior peso é associado a observação mais recente e o menor peso à observação mais antiga (HYNDMAN; ATHANASOPOULOS, 2018).

Uma das grandes vantagens dessa técnica é não precisar de variável externa para previsões, ela trabalha somente com dados e informações fornecidas pela própria série de dados. Além disso, esse método possui boa precisão, simples formulação e ajustes simples. Quando se trabalha com séries temporais de previsões de curto prazo, ela é bastante recomendada (VERÍSSIMO *et al.*, 2013).

A *Simple Exponential Smoothing* (SES), traduzido por Suavização Exponencial Simples, é o método mais simples de suavização e é propício a dados de previsão sem tendência ou padrão sazonal (HYNDMAN; ATHANASOPOULOS, 2018). Sua representação matemática é apresentada na Equação 2.3:

$$\hat{Y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \dots \quad (2.3)$$

Na Equação 2.3, a variação de pesos é controlada pelo parâmetro de suavização α , que varia de 0 a 1. Quando α estiver em um valor perto de 1, mais peso é atribuído as observações mais recentes, por sua vez, quando estiver mais perto de 0, menor é o peso atribuído. Há dois casos extremos: $\alpha = 0$ e $\alpha = 1$. No primeiro caso, $\alpha = 0$, a previsão dos futuros valores serão iguais a média dos dados históricos, conhecido como método da média. No segundo caso, $\alpha = 1$, todas as previsões serão definidas com o valor da última observação, conhecido como método *Naive* nas estatísticas (ZHANG, 2018).

2.5.5 *Random Walk*

O *Random Walk*, do português Caminho Aleatório, foi proposto por Karl Pearson em 1905, e consiste na sequência de movimentos aleatórios de um objeto matemático. Essa técnica pode ser aplicada em problemas como: caminho realizado por uma partícula subatômica ou caminho para um bêbado voltar para casa (RÊGO *et al.*, 2020). Pereira (2009) afirma que "o passeio aleatório é um protótipo para vários modelos estocásticos os quais as transições só

podem ser feitas entre vizinhos". Além disso, o passeio aleatório pode aparecer em uma, duas ou três dimensões e deve-se apresentar conforme a melhor condição.

O passeio aleatório em 1-D começa em sua origem ($y = 0$) e é movido em todas as etapas com igual probabilidade. O passeio aleatório em 2-D inicia na origem ($x = 0, y = 0$) e passos aleatórios em cada direção. Por exemplo, caso $(\Delta x, \Delta y) \in (-1, 0, 1)$ serão possíveis 9(nove) direções para o movimento em cada passo. E o passeio aleatório em 3-D, inicia na origem ($x = 0, y = 0, z = 0$), e nesse caso para $(\Delta x, \Delta y) \in (-1, 0, 1)$ serão possíveis 27(vinte e sete) direções para o movimento em cada passo. Tem-se, para exemplo de passeio aleatório no espaço 3D, um corpo se movendo em um volume (NILESH, 2018).

2.6 Métricas de Validação de Erros

Observando a qualidade das previsões, para verificar se o modelo tem boa assertividade, deve-se fazer a medição do erro, onde é verificado se os valores previstos estão próximo aos dos dados reais(RABELO, 2019). A seguir estão as principais métricas usadas para avaliar modelos de séries temporais neste trabalho:

2.6.1 Mean Absolute Error

O Erro Médio Absoluto, do inglês MAE, consiste em calcular o residual de cada ponto, no qual os valores residuais positivos e negativos não se anulam. Após esse agrupamento, calculamos então a média desses residuais (SILVA, 2019). Como segue na Equação 2.4.

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_{real} - Y_{previsto}| \quad (2.4)$$

onde:

n: número de dados utilizados no processo de estimativas,

Y real: valor real da amostra,

Y predito: o valor estimado na predição.

2.6.2 Root Mean Squared Error

A RMSE (do inglês, *Root Mean Squared Error (RMSE)* ou *Root Mean Square Deviation (RMSD)*), baseia-se na no quadrado da média das diferenças entre as previsões e observações reais. Pode-se assemelhar esta fórmula com a distância Euclidiana entre o vetor de

valores reais e o vetor de valores pressupostos, mediados pela quantidade de número de pontos 'n' (SILVA, 2019). Como segue na Equação 2.5.

$$RMSE = \sqrt{\sum_{t=1}^n \frac{(Y_{real} - Y_{previsto})^2}{n}} \quad (2.5)$$

onde:

n: número de dados utilizados no processo de estimativas,

Y real: valor real da amostra,

Y predito: o valor estimado na predição.

3 TRABALHOS RELACIONADOS

Em Kuplich *et al.* (2013) aplicou-se séries temporais para identificar a dinâmica fenológica das principais tipologias vegetais do Rio Grande do Sul, entre 2000 a 2010. Utilizando dados de *Enhanced Vegetation Index*, através da Transformada de Wavelet. O reconhecimento da fenologia em ciclos ou padrões sazonais em séries temporais, permite a observação de anomalias e os efeitos de mudanças climáticas ou ambientais. Em Gonzaga (2017), para detecção de anomalias em redes de sensores, foi abordada a elaboração de um *framework*. Isto porquê, essas redes estão sujeitas a leituras corrompidas e anômalas devido a restrições de processamento, comunicação e energia.

Outro trabalho relevante nesse estudo foi o de CAMPOS *et al.* (2006), que teve como propósito desenvolver um modelo de previsões para o preço do frango inteiro resfriado no grande atacado do estado de São Paulo, utilizando a metodologia ARIMA. Os dados históricos, organizados de forma mensal, eram entre os anos de 1996 a 2005 para predição de 2004 à 2005. Por fim, foi utilizada a métrica de erro *Mean Absolute Percentage Error* (MAPE) que obteve os melhores resultados entre 4% e 6%. Em MARTIN *et al.* (2016) também utilizou a metodologia ARIMA. O intuito do trabalho era prever demanda ou crescimento do número de automóveis no município de Joinville/SC. Além dessa metodologia, também foi abordado o modelo de suavização exponencial e foram utilizados dados de janeiro de 2003 à março de 2014. Utilizou-se como principais métricas de erro: ME(Erro Médio), MSE, MAE, MAPE.

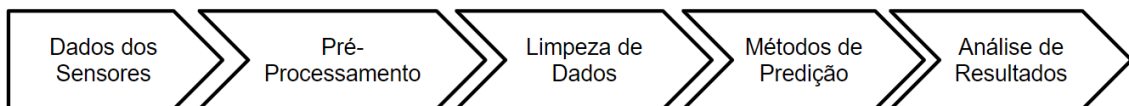
As séries temporais podem se apresentar de forma multidisciplinar. Para mostrar isso Otto *et al.* (2017) publica sua monografia na área de administração. O objetivo do seu trabalho é analisar modelos de suavização exponencial empregados na previsão de demanda das pequenas e médias empresas e seu potencial na gestão. Aplicou-se os modelos: suavização exponencial como um processo constante, modelo de Holt, modelos de Holt-Winters e, em seguida, comparou os modelos entre si, para verificar as vantagens e desvantagens no uso de cada um. Por fim, para auxiliar a decisão do gestor de empresa, foi proposto um *framework* quanto a escolha de qual modelo de suavização exponencial utilizar e outros três *frameworks* para auxiliar no cálculo da previsão de demanda para cada um desses modelos.

Por fim, em Veríssimo *et al.* (2013) é aplicado o método de suavização exponencial *Holt-Winters*, em uma empresa do setor mecânico, localizada na região norte do Estado de Santa Catarina. O método proposto tem como finalidade o planejamento na demanda de vendas e para a tomada de decisão na busca contínua da gestão otimizada do processo produtivo.

4 METODOLOGIA

No decorrer desta seção serão detalhadas as etapas para a metodologia proposta para a análise dos dados obtidos pelos sensores de trânsito. O procedimento metodológico da pesquisa é desenvolvido através de várias etapas, como será mostrado no fluxograma da Figura 9.

Figura 9 – Fluxograma da Metodologia.



Fonte – Elaborado pela Autora (2021).

A Figura 9 apresenta os passos necessários para a execução deste trabalho. A seguir, cada etapa será melhor detalhada.

4.1 Dados dos Sensores

A base de dados utilizada foi uma base privada do Projeto de Pesquisa e Desenvolvimento SINESP Big Data e Inteligência Artificial na Segurança Pública, desenvolvido pela Universidade Federal do Ceará (UFC) em parceria com o Ministério da Justiça e Segurança Pública. Ela contém 31.280 registros, onde os atributos são do tipo: *float64*, *object* e *int64* e é constituída por registros coletados no primeiro semestre de 2019. O desenvolvimento do trabalho deu-se através do *Google Colaboratory*, mais conhecido como *Colab*, que é um ambiente de execução em nuvem utilizando a linguagem de programação *Python*. Algumas das bibliotecas usadas foram: *pandas*, *numpy*, *seaborn*, *matplotlib* e *plotly* como principais, e outras de acordo com a necessidade.

4.2 Pré-Processamento de Dados

Nesta etapa, foi elaborada uma análise estatística para avaliar de forma geral quais sensores seriam considerados potencialmente com falhas. Sensores que apresentam comportamento padrão, são considerados sensores "saudáveis". Por sua vez, sensores que apresentam comportamento anormal, são considerados sensores "doentes". O objetivo da análise estatística é verificar quais sensores apresentam características anômalas em relação aos demais, e por

consequente, tentar descobrir causa desse acontecimento.

De início, foram verificadas as médias de velocidades registradas pelos sensores e seu equipamento(sensor) correspondente. Observou-se que alguns captavam imagens com velocidades muito baixas. Sensores que apresentavam velocidades menores que 10 km/h, são considerados sensores que não apresentam características padrão, isso porquê esse valor corresponde a uma velocidade de movimento muito inferior, logo nomeados de sensores com anomalia coletiva. Outra questão identificada neste trabalho para esse tipo de anomalia, são sensores que não se encontram no raio de localização da maioria dos dados da base e que apresentam um único registro durante toda a coleta para a criação da base de dados. Considerando que esses sensores encontrem-se dentro da cidade, existe o tráfego que torna inviável o registro de apenas uma ultrapassagem de veículos. Possivelmente, esses equipamentos não estavam registrando esses movimentos. Além disso, a quantidade de registros por sensores não acontece de maneira tão inferior, ou seja, não existe registros de duas ou três observações.

Para averiguar a saúde desses sensores individualmente, foram selecionados os dez sensores com a maior quantidade de registros, para analisar como esses se comportam ao longo do tempo. Observou-se que, existem equipamentos que apresentam alguns registros fora do padrão e um possível motivo seriam causas meteorológicas. Para isso, foi explorado no site da Fundação Cearense de Meteorologia e Recursos Hídricos (Funceme) a quantidade de chuva no dia correspondente a esse comportamento anômalo.

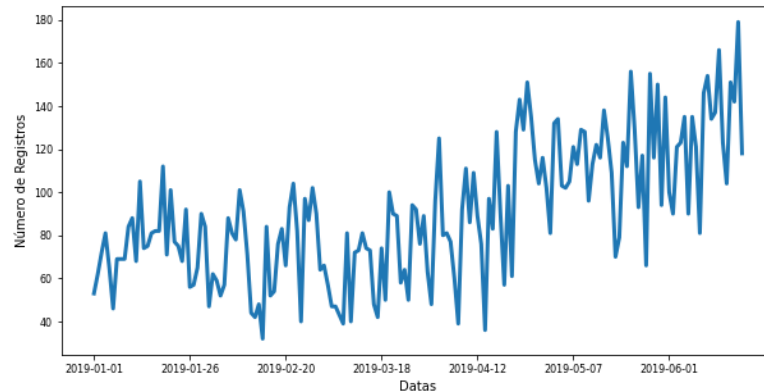
4.3 Limpeza de Dados

Neste trabalho, a limpeza de dados foi fundamental para um melhor resultado nas técnicas de predição. Durante a análise estatística do *dataset*, constatou-se que havia registro de dados que não era semelhantes com o comportamento da maioria. Dessa forma, optou-se pela remoção dos sensores anômalos com registros abaixo de 10 km/h, com apenas um registro e com localização fora do raio da região em estudo. Essa estratégia foi adotada pois o objetivo era analisar a qualidade dos sensores na região com maior frequência na base de dados. Além disso, os registros em outras localidades, eram registros únicos por região, que não trariam informações suficientes ou, até mesmo, registros que apresentavam velocidades nulas.

4.4 Métodos de Predição

À princípio, para a execução das técnicas de previsão, é necessário avaliar a estacionariedade da série. Nota-se, a plotagem da série na Figura 10.

Figura 10 – Não-Estacionariedade da Série



Fonte – Elaborado pela Autora (2021).

Observa-se, na Figura 10, que a série não apresenta comportamento constante, podendo sugerir que ela não é uma série estacionária. Um método bastante eficiente e utilizado para verificar a estacionariedade do *dataset* é o teste ADF. Na Figura 11, é mostrado o resultado desse teste na série estudada.

Figura 11 – Aplicação do Teste ADF na Série.

```

Test Statistic: -0.773195
p-value: 0.826869
Critical Values:
  1%: -3.471
  5%: -2.879
 10%: -2.576
Failed to Reject Ho - Time Series is Non-Stationary

```

Fonte – Elaborado pela Autora (2021).

É possível notar, através da Figura 11, que de fato, a série não é estacionária, pois seu “*Test Statistic*”(Estatística de Teste) está maior que o números de “*Critical Values*”(Valores Críticos). Quando isso acontece, é necessário aplicar técnicas para tornar essa série estacionária. No entanto, não existe uma técnica específica ou uma técnica universal para todos os modelos. Deve-se fazer diferentes tentativas para analisar qual se adapta melhor. Neste trabalho utilizou-se a técnica de transformação usando logaritmo, para diminuir o efeito da variância. Isso porquê, quando a variância não é constante, a previsão tende a ter erros maiores em determinados períodos. Na Figura 12, é mostrado o resultado do teste *Dickey-Fuller* após a transformação.

Figura 12 – Aplicação do Teste ADF após Transformação

```

Results of Dickey-Fuller Test:
Test Statistic           -6.373441e+00
Number of Observations Used  2.314693e-08
Critical Value (1%)      -3.472431e+00
Critical Value (5%)      -2.880013e+00
Critical Value (10%)     -2.576619e+00
  
```

Fonte – Elaborado pela Autora (2021).

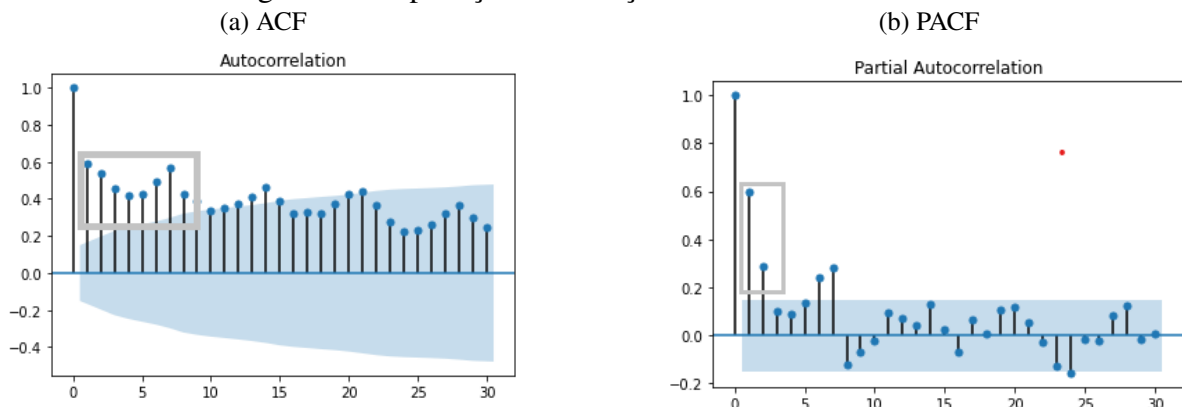
Através da Figura 12, observa-se que os valores críticos estão maiores que o valor de estatística de teste, tornando assim a série estacionária.

Após a estacionarização da série, é necessário separar os dados de treino e teste. Considerou-se dados de treino todos os registros de meses com exceção dos dois últimos, por sua vez, tornando os dois últimos meses os dados utilizados para teste. Além disso, também foram utilizadas outras formas de divisão para averiguar se as métricas de erros seguiam a mesma média. Primeiro, os dados de treino e teste foram divididos de maneira aleatória. Em seguida, foi aplicada uma técnica de soma sequencial para os dados de teste. Por fim, observou-se que as métricas seguiam a mesma média de erros.

Após as etapas anteriores, os algoritmos podem ser executados.

1. Arima: para execução deste algoritmo, é necessário as funções PACF e ACF para analisar a quantidade certa de parâmetros a ser utilizada, mostrados na Figura 13.

Figura 13 – Aplicação das Funções ACF e PACF na série.



Fonte – Elaborado pela Autora (2021).

Nos gráficos apresentados da Figura 13, tem-se que $p=2$, $d=1$ e $q=8$, formando o modelo $(2,1,8)$. No entanto, também foram executados outros valores para os parâmetros do método, a fim de identificar o melhor resultado. Dessa forma, foram escolhidos de forma aleatória os valores: $(1,1,1)$, $(1,0,1)$, $(1,0,0)$, $(0,0,1)$ e $(2,1,3)$.

2. *Simple Exponential Smoothing*: para esse algoritmo são usadas médias ponderadas e esses pesos são definidas pelo parâmetro de suavização α . Neste trabalho, utilizou-se três variantes de suavização exponencial simples. Foi aplicado $\alpha = 0,2$, $\alpha = 0,7$ e na terceira variação, usou-se a otimização automática que permite que os modelos de estatísticas encontrem automaticamente um valor otimizado, gerando $\alpha = 0,1553$.
3. *Prophet*: o desenvolvimento deste algoritmo, é dado através de sua biblioteca *fbprophet*. De início, é necessário verificar o nome das colunas da base de dado. Isto porquê, por padrão, a entrada é sempre um *dataframe* com duas colunas: *ds* e *y*. A *ds* é marcada pelos dados temporais e a coluna *y* refere-se aos dados quantitativos que serão preditos. A variável *ds* foi executada de duas maneiras. Na primeira execução é marcada pelo número de registros por dia e a segunda execução é marcada pelo número de registros no intervalo de hora por dia. Taylor e Letham (2018) afirma a importância de ressaltar que embora o *Prophet* seja otimizado, ele também pode ser projetado para ter parâmetros que podem ser ajustados sem conhecer os detalhes do modelo subjacente. No desenvolvimento desta pesquisa, utilizou-se parâmetros padrões. Posteriormente, é necessário instanciar e treinar os dados.
4. *Random Walk*: optou-se por trabalhar com o passeio aleatório simples unidimensional(1D), embora exista o modelo 2D e 3D, mas que não se adequaram ao modelo. De maneira geral, existe um ponto de partida que começa em 0 e em cada etapa se move com igual probabilidade.
5. *Walk Forward*: esse método percorre todo o *dataframe* calculando a média entre um número de linhas sugerido para gerar o resultado seguinte. Esse número de linhas é ajustado conforme o número inteiro escolhido para a "janela". Por exemplo, supondo que a janela tenha sido 5, a média será feita entre os cinco primeiros elementos e o resultado dará o sexto registro. Dessa vez, o último número fará parte do cálculo da média, juntamente com os quatro anteriores e formará o próximo registro. Assim sucessivamente até o último valor. Neste trabalho, após tentativas e erros, optou-se por uma janela de tamanho três.

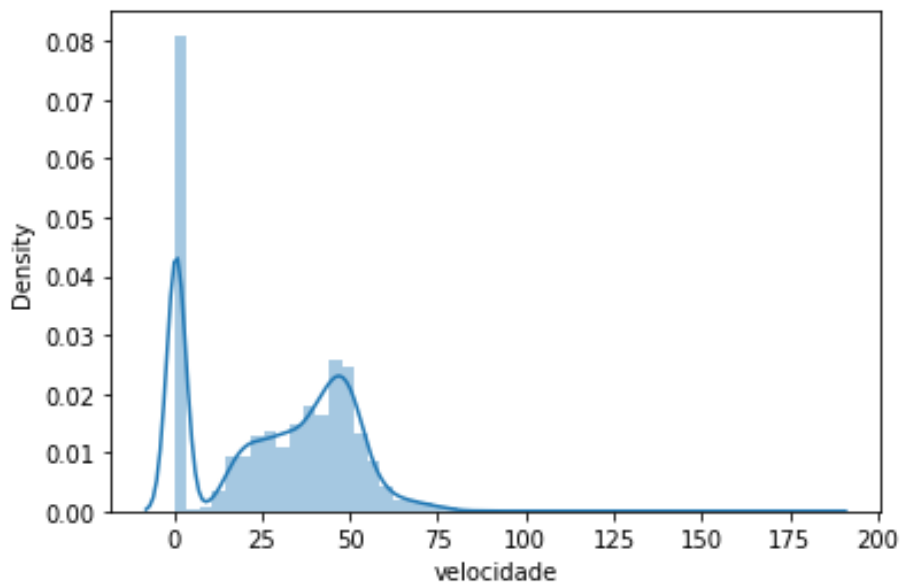
5 RESULTADOS

De acordo com a metodologia aplicada, neste capítulo serão expostos os resultados obtidos. Na Seção 5.1 serão mostrados os resultados referentes a análise sobre os sensores de trânsito. Na Seção 5.2 é mostrado os resultados relativos as técnicas de predição empregadas. Por fim, na Seção 5.3 será comparado os resultados.

5.1 Análise de Sensores

Como destacado anteriormente, a base de dados é formada por 31280 registros, 946 tipos de sensores e 135 observações do tipo *NULL*. Não apresentando registros duplicados. A Figura 14 mostra a distribuição da frequência das velocidades em geral registradas.

Figura 14 – Distribuição de frequência das velocidades registradas.



Fonte – Elaborado pela Autora (2021).

Na Figura 14, percebe-se a grande quantidade de registros com velocidades igual a 0 km/h. Logo, a necessidade de avaliar esse comportamento e verificar a saúde dos sensores. Para verificar a saúde dos sensores, algumas características não padrão foram consideradas anomalias que tornariam os sensores "doentes". Entre elas, sensores que apresentam registro de velocidades menores que 10 km/h, são considerados sensores potencialmente com falhas.

Na Tabela 1 será mostrada a quantidade de registros que estão nesse cenário e um gráfico que representa seu comportamento.

Tabela 1 – Sensores Potencialmente com Falhas

Velocidade	Quantidade de Registros
0	3886
1	4534
2	706
3	110
4	10
5	6
6	10
7	6
8	14
9	24
10	34

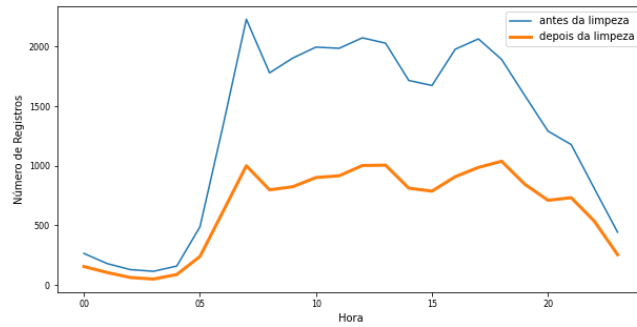
Fonte – Elaborada pela autora (2021)

Através da Tabela 1, nota-se que há 9340 registros com velocidades menores que 10 km/h. Conclui-se assim, que esses sensores apresentam anomalias coletivas, por ser uma característica incomum encontrada em um subconjunto de dados. Outra característica observada foi a respeito de sensores que não se encontravam no raio de localização que a grande maioria dos dados. Para isso, foi aplicado um filtro que eliminava os registros desses equipamentos. Dos 20 equipamentos com a maior quantidade de registros, 2(dois) estavam fora do raio de localização. Além disso, foi observado que esses equipamentos registravam velocidades nulas em sua pluralidade. Ademais, equipamentos que somavam apenas 1(um) registro durante todos os registros da base de dado, também foi considerado uma particularidade de anomalias coletivas. Esses sensores somavam em 58 equipamentos.

Por sua vez, para analisar comportamentos anômalos em sensores individuais, foram escolhidos os sensores com a maior quantidade de registros e observada seu comportamento ao longo do tempo. Em menos de 8% dos registros foram encontradas anomalias. Uma das causas encontradas, foi o fluxo de chuva muito alto que pode ser analisado através do site de meteorologia da Funceme. Chuvas acima de 70 mm são considerados chuvas fortes que podem comprometer o funcionamento do sensor. No entanto, não notou-se a necessidade de remoção desses registros, visto que apresentavam-se em mínimas quantidades. Além disso, esses sensores teriam informações que poderiam colaborar com a predição.

Após a análise do comportamento dos sensores e identificação dos sensores anômalos, efetuou-se a limpeza desses dados. Ao desenvolver essa etapa, percebeu-se mudanças nos registros em geral, em seus horários de pico e seus comportamentos, como será mostrado na Figura 15.

Figura 15 – Comparação dos registros antes e depois da limpeza de dados.



Fonte – Elaborado pela Autora (2021).

Na Figura 15, a quantidade de registros é somada de acordo com a hora, onde a linha azul representa os dados antes da limpeza e a linha laranja representa os dados após a limpeza. Antes de realizar esse procedimento, os picos com a maior quantidade de passagem de veículos se dava às 07h, 12h e 17h. Após a eliminação dos registros que não apresentavam informações suficientes para análise, constatou-se que os novos horários de picos seriam 12h, 13h e 18h.

5.2 Métodos de Predição

Assim, com os dados preparados para a predição, foram realizados a execução dos algoritmos. Onde sua qualidade e eficiência, foi avaliada através de métricas de erros, como mostrado a seguir:

1. ARIMA:

Durante a execução deste algoritmo, foram compiladas diferentes combinações de parâmetros a fim de comparar o melhor resultado. À princípio, testou os valores de parâmetros sugeridos pelas funções ACF e PACF, em seguida os valores escolhidos de forma aleatória. A seguir será apresentada, na Tabela 2, a comparação entre os parâmetros e cada uma das métricas de erro.

Tabela 2 – Comparação de Resultados do Método Arima

Parâmetros	RMSE	MAE
(2, 1, 8)	21.086	17.495
(2, 1, 3)	21.333	17.402
(1, 1, 1)	22.200	18.214
(1, 0, 0)	30.399	24.542
(1, 0, 1)	30.399	24.542
(0, 0, 1)	33.890	25.999

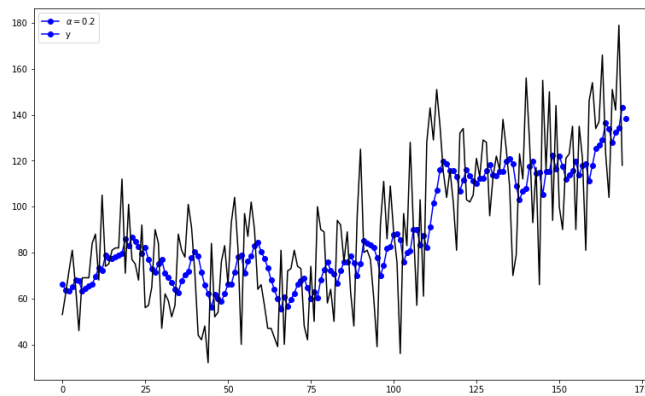
Fonte – Elaborada pela autora (2021).

Através da Tabela 2, observa-se que o melhor cenário foi para o parâmetro(2,1,8), valor sugerido inicialmente. Em suas variações, o modelo auto regressivo(1, 0, 0) comportou-se igualmente ao modelo auto regressivo com média móvel sem integração (1, 0, 1). Os demais modelos se comportaram de forma semelhante entre si.

2. *Simple Exponential Smoothing:*

De acordo com a metodologia da seção anterior, aplicou-se três variações para o parâmetro de suavização α , $\alpha = 0.2$, $\alpha = 0.7$ e $\alpha = 0.1553$, que é um valor automatizado. Serão expostos os resultados de erros obtidos a cada situação a seguir. Primeiramente, na Figura 16 , será mostrada a primeira visualização de variação, que será para $\alpha = 0.2$.

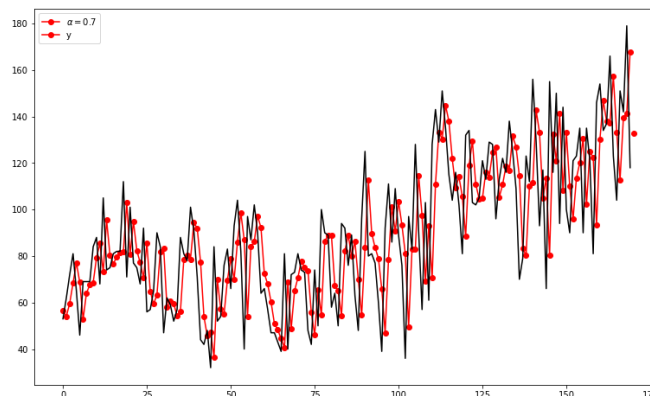
Figura 16 – SES com parâmetro de suavização $\alpha = 0.2$.



Fonte – Elaborado pela Autora (2021).

Na Figura 16, o eixo y representa a quantidade de registros e o eixo x representa o dia correspondente. A cor preta representa a evolução dos valores reais, já a cor azul representa os dados preditos pelo parâmetro de suavização escolhido. Como resultado, trazem: RMSE = 57.183 e MAE = 50.076. Na Figura 17, o gráfico para $\alpha = 0.7$.

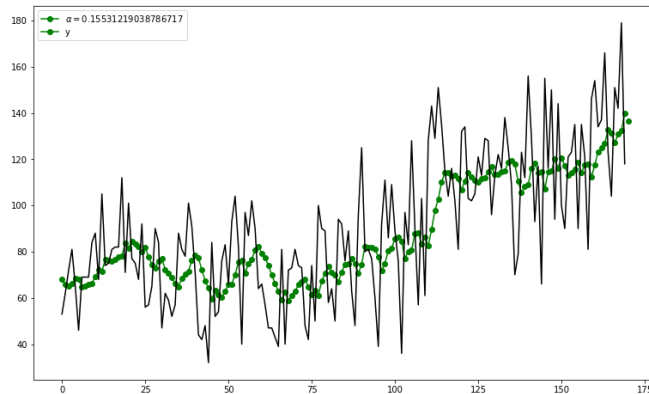
Figura 17 – SES com parâmetro de suavização $\alpha = 0.7$.



Fonte – Elaborado pela Autora (2021).

Na Figura 17, os eixos possuem a mesma representação da figura anterior. No entanto, a evolução dos valores reais também está sendo exibido pela cor preta, já a a cor vermelha estará representando os dados preditos pelo parâmetro de suavização escolhido, no caso $\alpha = 0.7$. Gerando como erro, RMSE = 52.809 e MAE = 45.725. Na Figura 18, o gráfico para α com valor automático.

Figura 18 – SES com parâmetro de suavização α automatizado.



Fonte – Elaborado pela Autora (2021).

Na última variação do parâmetro de suavização, na Figura 18, foi utilizado um α otimizado, que permite que os valores se adéquem automaticamente. Nesse caso, o valor gerado foi 0,1553. E, apresentado como métrica de erro, RMSE = 55.642 e MAE = 48.516.

3. Prophet:

Para esse algoritmo, foram utilizados duas maneiras diferentes para variável temporal. Na primeira execução, os dados foram divididos pela quantidade de registros por dia. E obteve como erro RMSE = 18.367 e MAE = 14.165. Na Figura 19, será representado o exemplo de sua saída.

Figura 19 – Resultado Saída *Prophet* com dados temporais diários.

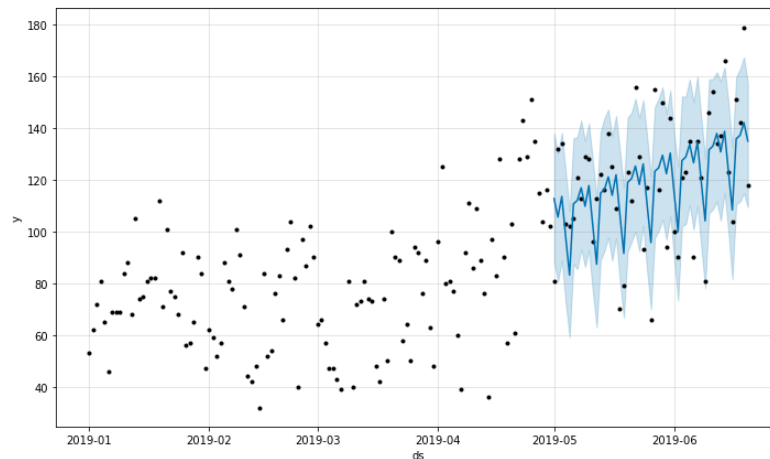
	ds	yhat	yhat_lower	yhat_upper
0	2019-05-01	112.775788	88.114906	138.012949
1	2019-05-02	105.647659	80.744320	130.397794
2	2019-05-03	113.611882	88.780781	138.222538
3	2019-05-04	98.112592	72.447384	123.212186
4	2019-05-05	83.183017	59.049259	107.903699

Fonte – Elaborado pela Autora (2021).

Na Figura 19, é apresentado a saída do algoritmo, onde a coluna ds é representado pelos

dados temporais e a coluna ao seu lado, chamada de *yhat*, pelo valor predito. Logo em seguida, existem as colunas *yhat lower* e *yhat upper*, que fazem referência aos seus valores inferiores e superiores respectivamente. Na Figura 20, será ilustrado visualmente como esses dados se comportam.

Figura 20 – Gráfico *Prophet* com dados temporais diários.



Fonte – Elaborado pela Autora (2021).

Na Figura 20, o eixo *y* representa a quantidade de registros e o eixo *x*, atendendo por *ds*, representa os dados temporais, que são os dias respectivos. No gráfico são exibidos formas pontilhadas pretas e uma linha azul. Os pontilhados são os dados reais, já a linha azul representa os dados preditos.

Na segunda execução, os dados foram divididos pela quantidade de registros por intervalos de hora durante os dias respectivos. Ou seja, os registros são contados todos os dias com o intervalo de uma hora. E obtive como erro $RMSE = 3.305$ e $MAE = 2.533$. Para melhor explicação da saída desses dados nessa forma temporal, será exibida a Figura 21.

Figura 21 – Resultado Saída *Prophet* com dados temporais por hora diariamente.

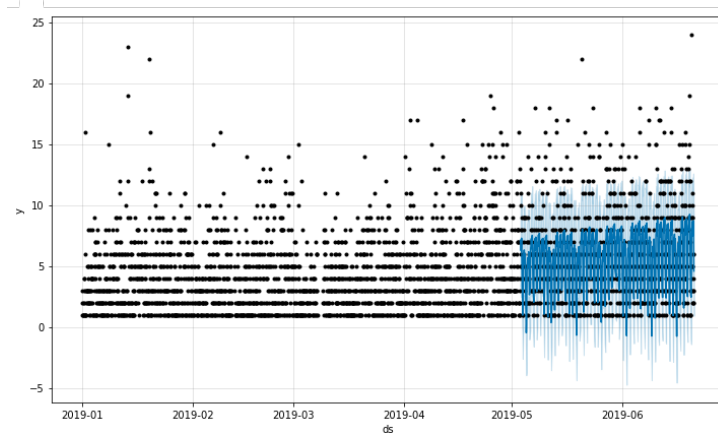
	<i>ds</i>	<i>yhat</i>	<i>yhat_lower</i>	<i>yhat_upper</i>
0	2019-05-03 11:00:00	6.634472	2.693781	10.394851
1	2019-05-03 12:00:00	6.822563	3.194986	10.568556
2	2019-05-03 13:00:00	6.691516	3.041830	10.246182
3	2019-05-03 14:00:00	6.327034	2.753394	10.153291
4	2019-05-03 15:00:00	6.100964	2.553396	9.783786

Fonte – Elaborado pela Autora (2021).

Anteriormente, na Figura 21, por sua vez, a coluna *ds* representa dados temporais divididos

por hora diariamente. Logo em seguida são exibidos os valores preditos e seu intervalo de variação. Dessa forma, na Figura 22, é ilustrado como esses dados se comportam.

Figura 22 – Gráfico *Prophet* com dados temporais por hora diariamente.



Fonte – Elaborado pela Autora (2021).

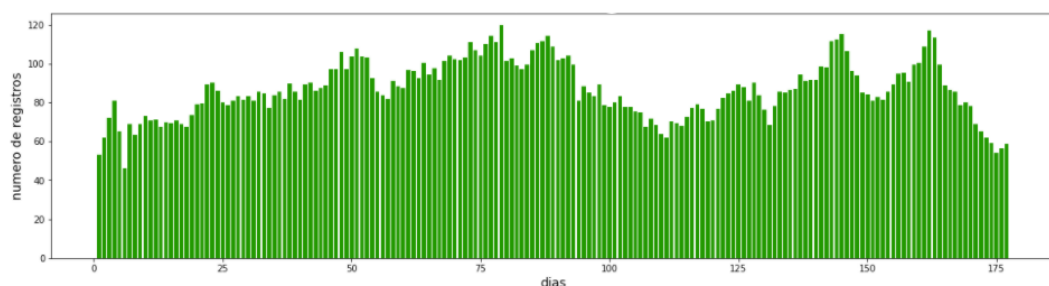
Na Figura 22, o eixo y representa a quantidade de registros e o eixo x, atendendo por ds, representa o intervalo de hora pelos dias respectivos. Os pontilhados são os dados reais, já a linha azul representa os dados preditos.

O *Prophet* produz previsões confiáveis, de maneira automática e rápida, sem muito esforço manual. Ele é robusto para *outliers*, dados ausentes e mudanças significativas em sua série temporal (TAYLOR; LETHAM, 2018). Devido a essas características, a técnica se adequou bem ao modelo.

4. *Random Walk*:

Nesse método, o caminho aleatório unidimensional, apresentou como resultado das métricas de erro RMSE = 56.072 e MAE = 43.415. Sua representação gráfica, é mostrada na Figura 23.

Figura 23 – Gráfico *Random Walk 1D*.



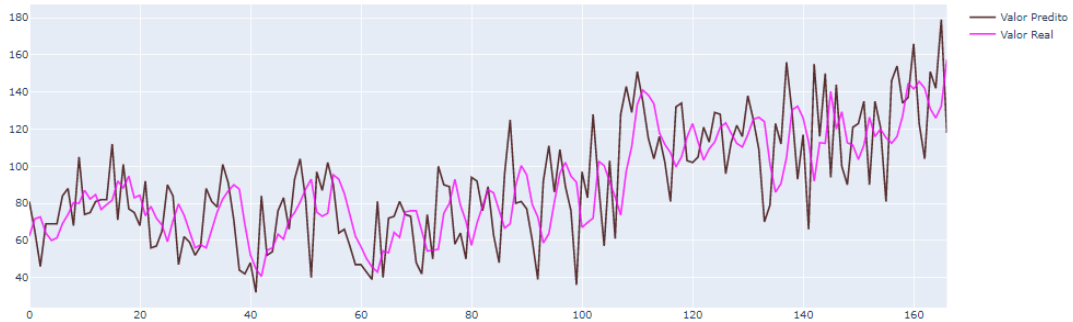
Fonte – Elaborado pela Autora (2021).

Na Figura 23 é mostrado como os dados se comportam através da execução deste método.

5. *Walk Forward*:

Nesse algoritmo, são aplicadas médias móveis como o número de janela, neste trabalhou usou-se janela de tamanho igual a 3. Têm-se, como resultado de métricas de erro, RMSE = 24.722 e MAE = 20.238. Sua representação gráfica, é mostrada na Figura 24.

Figura 24 – Gráfico *Walk Forward*.



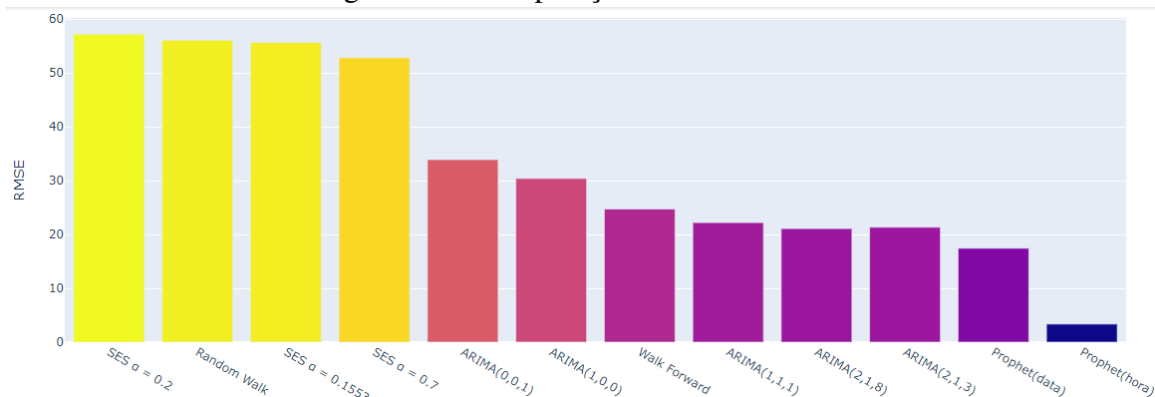
Fonte – Elaborado pela Autora (2021).

Na Figura 24, é mostrado a comparação entre os dados reais e os dados preditos, dados reais simbolizados pela cor lilás e valores preditos pela cor roxa. O gráfico é desenvolvido através da biblioteca *plotly*.

5.3 Análise de Resultados

Após acompanhar a execução de todas as técnicas aplicadas, será mostrado nas Figuras 25 e 26, a comparação dos resultados obtidos em cada métrica de maneira visual.

Figura 25 – Comparação resultado RMSE.



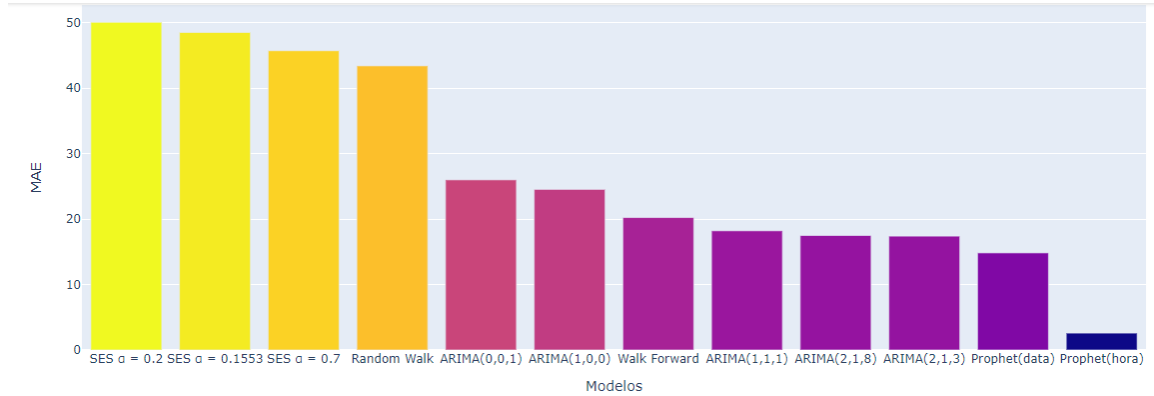
Fonte – Elaborado pela Autora (2021).

A Figura 25 apresenta os resultados referentes a métrica de erro RMSE. Através da

visualização, nota-se, que em suas duas execuções, o *Prophet* apresentou taxa de erro abaixo dos demais métodos.

A seguir, na Figura 26, será exposto a comparação entre o MAE.

Figura 26 – Comparação resultado MAE.



Fonte – Elaborado pela Autora (2021).

A Figura 26 apresenta os resultados referente a métrica de erro MAE. Quando comparada as métricas anteriores, foi a de menor taxa de erro e, também, teve o *Prophet* com o melhor desempenho.

Por fim, na Tabela 3, será mostrado uma tabela para avaliação geral e comparação entre os resultados de erros obtidos nos métodos aplicados. Nas técnicas ARIMA e SES, que possuíam diferentes variações de parâmetros, foi exibido somente o melhor valor revelado entre elas.

Tabela 3 – Comparação de Resultados dos Métodos de Predição e Métricas de Erro

Modelos	RMSE	MAE
Prophet (com intervalo de hora)	3.359	2.587
Prophet (com intervalo de dias)	18.367	14.839
Arima (2, 1, 8)	21.086	17.495
Walk Forward	24.722	20.238
Expon. Smoothing ($\alpha=0.7$)	52.809	45.725
Random Walk	56.072	43.415

Fonte – Elaborada pela autora (2021)

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho foi desenvolvido com o objetivo de analisar a qualidade dos dados obtidos pelos sensores de trânsito através de séries temporais. Dessa forma, através do estudo da literatura, viu-se a necessidade de analisar e reconhecer padrões anômalos nessas informações capturadas pelos equipamentos e testar técnicas de séries temporais.

Identificou-se a existência de características anômalas em vários conjuntos de dados. Apontou-se com essas características, sensores que capturavam apenas um registro de ultrapassagem, sensores que registravam velocidades inferiores à 10 km/h e registros fora do raio de localização da base de dados. Através dessa observação, foi possível notar que estas anomalias não se apresentam de forma individual, ou seja, um conjunto de dados pode apresentar comportamento anômalo semelhante a outro conjunto de dados. No caso deste trabalho, geralmente, equipamentos que apontavam apenas um registro ou equipamentos fora do raio de localização, também registravam velocidades inferiores a um limite possível ou velocidades nulas. Ou, até mesmo, sensores fora do raio de localização eram sensores que só apresentavam um registro.

Além disso, através das técnicas de predição: *Arima*, *Prophet*, *Exponential Smoothing*, *Walk Forward* e *Random Walk*, pode-se avaliar o comportamento futuro dos dados e como eles se adequavam aos modelos. Embora tenham sido aplicados modelos clássicos de séries temporais, pode-se constatar que o método que apresentou melhor comportamento foi o da biblioteca do *Facebook*, conhecido como *Prophet*, com intervalo de hora diariamente. O modelo utiliza regressão aditiva e tem como componentes: tendência, sazonalidade e feriados. Ele obteve como erro, RMSE= 3.359 e MAE=2.587, que significa que os dados preditos apresentam esse valor para mais ou para menos comparados com os valores reais. Outro fato, é que a métrica de erro MAE apresentou resultado inferior as demais métricas em todos os modelos de predição.

Em suma, pode-se concluir, como resultado do trabalho, que os objetivos levantados foram alcançados. Através do estudo, foram identificados os padrões anômalos em alguns equipamentos, sendo importante para auxiliar os agentes de trânsito a detectá-los para sua correção. Ademais, dentre os modelos de predição, pode analisar o comportamento futuro desses dados.

Como trabalhos futuros, sugere-se o uso de novas bases de dados para uma nova avaliação a cerca de possíveis análises do comportamento dos dados, reconhecimento de outras anomalias e a implementação de novas técnicas de predição de dados.

REFERÊNCIAS

- ABEETRANS. Big data onipresente, onisciente. Uma publicação da Associação Brasileira das Empresas de Engenharia de Trânsito, p. 1–3, 2016.
- ALVES, R. P.; ROSSI, R. G. Avaliação de técnicas de aprendizado de máquina indutivo supervisionado para classificação automática de textos em fluxo de notícias. Universidade Federal de Mato Grosso do Sul (UFMS), 2018.
- ARZAMENDIA, O. **An Introduction to Time-series Analysis Using Python and Pandas.** [S.l.], 2019. Dados Publicados por Medium. Disponível em: <<https://bit.ly/3t6Gwfc>>. Acesso em: 16 fev. 2021.
- BANDURA, E.; BUENO, J. C. M.; JADOSKI, G. S.; JUNIOR, G. F. R. Applications of the arima model for time series data analysis. **Applied Research & Agrotechnology**, v. 12, n. 03, p. 145–150, 2019.
- BECKER, M. H. Modelos para previsão em séries temporais: uma aplicação para a taxa de desemprego na região metropolitana de porto alegre. Porto Alegre, 2010.
- BEZERRA, M. I. S. Apostila de análise de séries temporais. **Curso de**, 2006.
- BROWNLEE, J. **How to Check if Time Series Data is Stationary with Python.** [S.l.], 2016. Dados Publicados por Machine Learning Mastery. Disponível em: <<https://machinelearningmastery.com/time-series-data-stationary-python/>>. Acesso em: 16 fev. 2021.
- CAMPOS, P. A. C.; CLEMENTE, A.; CORDEIRO, A. A. L. D. Aplicação do modelo arima para previsão do preço do frango inteiro resfriado no grande atacado do estado de são paulo. In: **Anais do Congresso Brasileiro de Custos-ABC.** [S.l.: s.n.], 2006.
- CARVALHO, H. V. de; CARVALHO, E. C.; ARRUDA, H.; IMPERATRIZ-FONSECA, V.; SOUZA, P. de; PESSIN, G. Detecção de anomalias em comportamento de abelhas utilizando redes neurais recorrentes. In: SBC. **Anais do IX Workshop de Computação Aplicada a Gestão do Meio Ambiente e Recursos Naturais.** [S.l.], 2018.
- CHINO, D. Y. T. **Mineração de padrões frequentes em séries temporais para apoio à tomada de decisão em agrometeorologia.** Tese (Doutorado) — Universidade de São Paulo, 2014.
- COUTINHO, T. **11 bibliotecas Python aplicadas na Ciência de Dados.** [S.l.], 2020. Dados publicados pelo Grupo Voitto. Disponível em: <<https://www.voitto.com.br/blog/artigo/principais-bibliotecas-python>>. Acesso em: 24 fev. 2021.
- EHLERS, R. S. Análise de séries temporais. **Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná**, 2007.
- FIGUEIRÊDO, I. S.; GUARIEIRO, L.; SANTOS, A.; NASCIMENTO, E. Algoritmo de aprendizagem de máquina não supervisionado para detecção de anomalias em séries temporais multivariadas aplicadas a temporadas de furacões. **V Seminário de Avaliação de Pesquisa Científica e Tecnológica [Internet]. Salvador, Bahia: SENAI CIMATEC**, p. 3, 2020.

- FREITAS, I. W. S. d. *et al.* Um estudo comparativo de técnicas de detecção de outliers no contexto de classificação de dados. Universidade Federal Rural do Semi-Árido, 2019.
- GOMES, P. C. T. **Etapas do Pré-Processamento de dados**. [S.l.], 2019. Dados Publicados por Data Geeks. Disponível em: <<https://www.datageeks.com.br/pre-processamento-de-dados/>>. Acesso em: 16 fev. 2021.
- GONZAGA, M. d. S. Implementação de um framework para a detecção contextual de anomalias em dados de sensores. 2017.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. [S.l.]: OTexts, 2018.
- KUPLICH, T. M.; MOREIRA, A.; FONTANA, D. C. Série temporal de índice de vegetação sobre diferentes tipologias vegetais no rio grande do sul. **Revista Brasileira de Engenharia Agrícola e Ambiental**, SciELO Brasil, v. 17, n. 10, p. 1116–1123, 2013.
- LYLA, Y. **Um Início Rápido da Previsão de Séries Temporais com um Exemplo Prático usando o FB Prophet**. 2019.
- MARTIN, A. C.; HENNING, E.; WALTER, O. M. F. C.; KONRATH, A. C. Análise de séries temporais para previsão da evolução do número de automóveis no município de joinville. **Revista ESPACIOS| Vol. 37 (Nº 06) Año 2016**, 2016.
- MATA, F. F. D. G. da. **Investigando Métodos Inteligentes para Detecção de Anomalias em Comportamento de Insetos Sociais**. Dissertação (Mestrado) — UNIVERSIDADE FEDERAL DO PARÁ, 2017.
- MIGON, H. Análise de séries temporais. **Departamento de Métodos Estatísticos, Instituto de Matemática, UFRJ, Disponível em: <http://acd.ufrj.br/~dani/pdf/slidespartefrequentista.pdf> Acesso em, v. 4, 2007.**
- MING, S. H. **Fiscalização eletrônica de trânsito**. 2006.
- NILESH. **Data Science. Simulate Random Walks With Python**. [S.l.], 2018. Dados publicados por Towards Data Science. Disponível em: <<https://bit.ly/3esYyUT>>. Acesso em: 15 fev. 2021.
- NOVA FILHO, V.; PESSOA, S. R. **Previsão de séries temporais utilizando pools de preditores criados a partir do particionamento da série e da divisão da tarefa de previsão**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2015.
- OLIVEIRA, B. **CARACTERÍSTICAS DAS SÉRIES TEMPORAIS**. [S.l.], 2019. Dados publicados pelo blog DATA SCIENCE, PRODUÇÃO CIENTÍFICA. Disponível em: <<https://operdata.com.br/blog/caracteristicas-das-series-temporais/>>. Acesso em: 12 jan. 2021.
- OLIVEIRA, M. **As 30 melhores bibliotecas e pacotes Python para Iniciantes**. [S.l.], 2019. Originalmente escrito por UbuntuPit com modificações e melhorias por Terminal Root. Disponível em: <<https://bit.ly/2PL4ttW>>. Acesso em: 24 fev. 2021.
- OTTO, L. *et al.* Uma análise de modelos de suavização exponencial e sua aplicabilidade à previsão de demanda na ótica da administração. Florianópolis, SC, 2017.
- PEREIRA, F. A. C. Passeio aleatório. 2009.

- PROPHET. **Forecasting at scale**. [S.l.], 2017. Disponível em: <<https://facebook.github.io/prophet/>>. Acesso em: 18 jan. 2021.
- RABELO, L. **Princípios básicos para criar previsões de Séries Temporais**. [S.l.], 2019. Dados Publicados por Medium. Disponível em: <<https://bit.ly/30xgYLN>>. Acesso em: 15 jan. 2021.
- RÊGO, M. S.; FIGUEIREDO, G. M.; BRITO, F. W. P.; BARROS, P. H. M.; TRINDADE, N. d. J. da S.; GOMES, R.; SILVA, J. dos S.; OLIVEIRA, L. V.; PORTO, J. A. S. Simulação de integração numérica: O problema “random walk” e o método de monte carlo. 2020.
- REIS, M. M. **INE 7001 Análise de Séries Temporais**. [S.l.], 2015. Universidade Federal de Santa Catarina. Disponível em: <<https://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf>>. Acesso em: 18 fev. 2021.
- REVISTA SIM. **Fiscalização inteligente**. [S.l.], 2018. Disponível em: <<http://patrocinados.estadao.com.br/arteris/fiscalizacao-inteligente/>>. Acesso em: 18 jan. 2021.
- SANTANA, R. **Previendo Dados com Janelas Deslizantes usando Python**. [S.l.], 2020. Dados publicados por Minerando Dados. Disponível em: <<https://minerandodados.com.br/prevendo-dados-com-janelas-deslizantes-usando-python/>>. Acesso em: 15 fev. 2021.
- SANTOS, M. G. dos; Lins Júnior, P. R. Avaliação de técnicas estatísticas para detecção de anomalias aplicadas à internet das coisas. **Revista de Tecnologia da Informação e Comunicação**, v. 8, n. 1, p. 13–18, 2018.
- SILVA, D. H. E. da. **ML - Métricas de Regressão**. 2019. Disponível em: <<https://bit.ly/30IV88t>>. Acesso em: 16 jan. 2021.
- SIQUEIRA, F. **Radares com falhas fazem com que 70% das multas de trânsito seja invalidado**. [S.l.], 2015. Disponível em: <<https://bit.ly/3bzmmo8>>. Acesso em: 03 fev. 2021.
- SOUZA, C. E. **Séries temporais com Machine Learning — Parte 1**. [S.l.], 2020. Dados publicados pelo blog oficial da comunidade Data Hacker. Disponível em: <<https://bit.ly/3vbAOuf>>. Acesso em: 12 jan. 2021.
- SOUZA, C. M. G. de. **Modelos de previsão de séries temporais aplicados ao setor editorial brasileiro**. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2017.
- SOUZA, R. Metodologias para a análise e previsão de séries temporais univariadas e multivariadas. **Brazilian Review of Econometrics**, v. 1, n. 2, p. 78–105, 1981.
- TAYLOR, S. J.; LETHAM, B. Forecasting at scale. **The American Statistician**, Taylor & Francis, v. 72, n. 1, p. 37–45, 2018.
- VASCONCELLOS, P. **Dicas para criar um modelo de previsão de séries temporais**. [S.l.], 2018. Dados Publicados por Medium. Disponível em: <<https://bit.ly/3byNmEn>>. Acesso em: 16 fev. 2021.
- VERÍSSIMO, A. J.; ALVES, C. da C.; HENNING, E.; AMARAL, C. E. do; CRUZ, A. C. da. Métodos estatísticos de suavização exponencial holt-winters para previsão de demanda em uma empresa do setor metal mecânico. **Revista Gestão Industrial**, v. 8, n. 4, 2013.

WERNER, L.; RIBEIRO, J. L. D. Previsão de demanda: uma aplicação dos modelos box-jenkins na área de assistência técnica de computadores pessoais. **Gestão & Produção**, SciELO Brasil, v. 10, n. 1, p. 47–67, 2003.

ZHANG, A. **How to Build Exponential Smoothing Models Using Python: Simple Exponential Smoothing, Holt, and Holt-Winters**. [S.l.], 2018. Dados publicados por Medium. Disponível em: <<https://bit.ly/3rzCtaP>>. Acesso em: 15 fev. 2021.