UNIVERSIDADE ESTADUAL PAULISTA - UNESP
Instituto de Ciências e Engenharia - Câmpus de Itapeva

ALISSON ARAGÃO DOS SANTOS

PREVISÃO UNIVARIADA DE CHAMADAS EM UM CALL CENTER RECEPTIVO

Itapeva - SP
2024

ALISSON ARAGÃO DOS SANTOS

PREVISÃO UNIVARIADA DE CHAMADAS EM UM CALL CENTER RECEPTIVO

Trabalho de Conclusão de Curso (TCC) apresentado ao Conselho de Curso de Engenharia de Produção, da Universidade Estadual Paulista (Unesp), Instituto de Ciências e Engenharia, Câmpus de Itapeva, como parte dos requisitos para obtenção do diploma de Graduação em Engenharia de Produção.

Orientador: Prof. Dr. Carlos de Oliveira Affonso

Itapeva - SP
2024

S237p

Santos, Alisson Aragão dos
Previsão univariada de chamadas em um call center receptivo / Alisson Aragão dos Santos. -- Itapeva, 2024
98 p. : il., tabs.

Trabalho de conclusão de curso (Bacharelado - Engenharia de Produção) - Universidade Estadual Paulista (UNESP), Instituto de Ciências e Engenharia, Itapeva
Orientador: Carlos de Oliveira Affonso

1. Centros de atendimento ao cliente. 2. Inteligência artificial. 3. Previsão. 4. Analise de series temporais. 5. Machine learning. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Instituto de Ciências e Engenharia - Câmpus de Itapeva

ALISSON ARAGÃO DOS SANTOS

PREVISÃO UNIVARIADA DE CHAMADAS EM UM CALL CENTER RECEPTIVO

Trabalho de Conclusão de Curso para obtenção do título de Bacharel em Engenharia de Produção, da Universidade Estadual Paulista - UNESP - Campus de Itapeva.

BANCA EXAMINADORA

Orientador: Prof. Dr. Carlos de Oliveira Affonso
Universidade Estadual Paulista - UNESP - Campus de Itapeva.

2º Examinador: Prof. Dr. Antônio Francisco Savi
Universidade Estadual Paulista - UNESP - Campus de Itapeva.

3º Examinador: Prof. Dr. Marcel Yuzo Kondo
Universidade Estadual Paulista - UNESP - Campus de Itapeva.

Itapeva, 14/11/2024.

Dedico o trabalho primeiramente a

Deus e a minha família.

AGRADECIMENTOS

Agradeço a Deus, que me guiou durante toda a graduação e deu força para continuar em todos os momentos em que pensei que não conseguiria.

À minha mãe, Aparecida, que sempre confiou em mim e esteve ao meu lado em todas as decisões. Sua dedicação incondicional e disposição para abdicar de tudo pelo bem-estar meu e dos meus irmãos. Agradeço por todo o amor e cuidado que sempre nos ofereceu.

Ao meu pai, Izaelcio, cuja trajetória de vida é marcada por luta e renúncia. Ele sempre renunciou a seu próprio conforto para me proporcionar um futuro melhor, nunca deixando de me incentivar a estudar e a seguir em frente. Sua força e generosidade foram e continuam sendo exemplos que levarei para toda a vida.

Aos meus irmãos, Igor e Leonardo, pela parceria em momentos que guardarei para resto da minha vida e por estarem sempre presentes quando precisei.

À minha esposa, Cintia, o amor da minha vida e minha companheira constante, especialmente nesta reta final da minha formação. Seu apoio, compreensão, e os conselhos que sempre acalmam meu coração foram fundamentais.

A meus avós paternos, Maria Helena e Orlando, que foram como segundos pais para mim. Agradeço a eles por todo o apoio e amor ao longo da minha vida.

Aos meus tios, João e Iraci, que foram para mim como avós maternos. Agradeço o carinho, apoio e presença ao longo da minha vida.

À minha sogra, que é como uma mãe para mim. Agradeço a ela pelo carinho, apoio e por ter me acolhido com tanto amor em sua família.

À toda minha família, que de forma direta ou indireta participaram da minha jornada.

Ao meu orientador, Prof. Carlos, pela oportunidade, confiança e paciência, e por todo o suporte durante a realização deste trabalho.

Aos meus colegas de graduação, que fizeram desta jornada uma experiência única e significativa, através de estudos em grupo, conversas motivadoras, ajuda mútua e parcerias valiosas.

Agradeço também a todos os professores que contribuíram para minha formação, cuja essência e ensinamentos levarei comigo para sempre. Estendo minha gratidão a todos os docentes que passaram pela minha vida, desde o ensino fundamental até o ensino médio.

Por fim, agradeço a todos que, de alguma forma, direta ou indireta, contribuíram para a realização deste trabalho.

“A mente que se abre a uma nova ideia
jamais voltará ao seu tamanho original”
(Albert Einstein).

RESUMO

A experiência em operações de *call center* e atendimento revelou uma oportunidade promissora para a aplicação de um modelo capaz de prever o volume de ligações, utilizando conceitos de séries temporais. Este trabalho tem como objetivo investigar e desenvolver um método para a previsão de chamadas, buscando aprimorar a eficiência operacional e o planejamento estratégico no setor de um *call center* receptivo, por meio de estudos e aplicação prática de abordagens de modelagem estatística clássica e aprendizado de máquina. Para essa finalidade, foi realizada uma análise comparativa entre os modelos ARIMA e *Random Forest*, utilizando um conjunto de dados reais de um *call center* localizado em *Cincinnati, Ohio, EUA*. A base inicial de dados compreendia todos os registros de chamadas de 2014 a 2023, sendo transformada em uma tabela de volumetrias mensais para as previsões. Para o desenvolvimento das modelagens, foi utilizada a linguagem de programação Python, juntamente com bibliotecas de análise de dados, como *statsmodels* e *sklearn*. O fluxo de trabalho seguiu três etapas principais: transformação e análise exploratória dos dados, previsão com o modelo ARIMA e previsão com o modelo *Random Forest*. Os resultados obtidos indicaram que ambos os modelos apresentaram ajustes satisfatórios. Diferentemente do observado na literatura, nesta base de dados específica, o modelo ARIMA superou o *Random Forest* em métricas de acurácia. No entanto, observou-se visualmente que o *Random Forest* conseguiu capturar padrões nos dados que o modelo clássico não conseguiu. As conclusões destacam a importância de analisar o *trade-off* entre a complexidade do modelo e a precisão das previsões, o que depende do cenário de aplicação. Além disso, foi identificada a oportunidade de aprimoramento em ambas as abordagens, que podem ser fortalecidas com a inclusão de outras variáveis nos modelos para melhorar os ajustes aos dados.

Palavras-chave: previsão; *call center*; arima; *random forest.*

ABSTRACT

Experience in call center and customer service operations revealed a promising opportunity for applying a model capable of predicting the volume of calls, using time series concepts. This work aims to investigate and develop methods for call forecasting, seeking to improve operational efficiency and strategic planning in the inbound call center sector, through studies and practical application of classical statistical estimation and machine learning approaches. For this purpose, a comparative analysis between the ARIMA and Random Forest models was performed, using a real dataset from a call center located in Cincinnati, Ohio, USA. The initial database comprised all call records from 2014 to 2023, which were transformed into a monthly volume table for forecasting. To develop the models, it was used the Python programming language, with use of data analysis libraries, such as statsmodels and sklearn. The workflow followed three main steps: transformation and exploratory analysis of the data, forecasting with the ARIMA model, and forecasting with the Random Forest model. The results obtained indicated that both models presented satisfactory adjustments. Unlike what was observed in the literature, in this specific database, the ARIMA model outperformed the Random Forest model in accuracy metrics. However, it was visually observed that Random Forest was able to capture patterns in the data that the classical model could not. The conclusions highlight the importance of analyzing the trade-off between model complexity and forecast accuracy, which depends on the application scenario. In addition, opportunities for improvement were identified in both approaches, which can be improved through addition of the other variables in the models to the more accurate adjustments to the data.

Keywords: forecasting; call center; arima; random forest.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AdaBoost	Adaptive Boosting
ABRAREC	Associação Brasileira das Relações Empresa Cliente
BEMD	Associação Brasileira de Marketing Direto
ABT	Associação Brasileira de Telesserviços
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ACD	Automatic Call Distributor
ANI	Automatic Number Identification
ASA	Average Speed of Answer
Bagging	Bootstrap Aggregating
CAR	Call Answer Rate
CTI	Computer Telephone Integration
R^2	Coeficiente de Determinação
BIC	Critério de Informação Bayesiano
AIC	Critério de Informação de Akaike
CRM	Customer Relationship Management
DNIS	Dialed Number Identification Service
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
i.i.d.	Identicamente Distribuídas
I	Integrated
IVR	Interactive Voice Response
IA	Inteligência Artificial
MMS	Médias Móveis Simples
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Square Error
MA	Moving Average
PABX	Private Automatic Branch Exchange
PROBARE	Programa Brasileiro de Autorregulamentação

PSTN	Public Switched Telephone Network
RMSE	Root Mean Squared Error
DPS	Secretaria de Serviços Públicos
SL	Service Level
CSR	Solicitação de Atendimento ao Cidadão
SEH	Suavização Exponencial de Holt
SHW	Suavização Exponencial de Holt-Winters
SES	Suavização Exponencial Simples
SVM	Support Vector Machine
SVR	Support Vector Regression
ADF	Teste de Dickey-Fuller Aumentado
CSV	Arquivo Separado por Vírgulas
AM	Aprendizado de Máquina
XGBoost	Extreme Gradient Boosting
SAC	Serviço de Atendimento ao Cidadão

SUMÁRIO

1 INTRODUÇÃO

Atualmente, o termo *call center* é conhecido mundialmente pela maioria das pessoas, haja vista que, está presente em diversas situações de interações, como vendas, suporte e assistência técnica. O conceito de *call center* surgiu em 1880, após quatro anos da primeira patente do telefone. Nessa época, um vendedor de doces organizou uma operação com uma centena de pessoas que realizavam vendas e prospecção de clientes a partir de uma lista de contatos telefônicos (Mancini, 2006).

O *call center* pode ser definido como um local de alta concentração de pessoas e tecnologias de telefonia, com proposito de oferecer serviços relacionados a vendas, marketing, cobranças e suporte ao cliente. Ao decorrer do tempo, agregou-se vários outros significados, em decorrência das diversas mudanças desde seu surgimento. Atualmente, o *call center* pode ser chamado de central de atendimento, *telemarketing* ou *televendas* e *contact center* (Dawson, 1998).

As várias denominações de *call centers* são caracterizadas pela segmentação das atividades desenvolvidas e tecnologias utilizadas. O termo centrais de atendimento é mais abrangente, contemplando todas as atividades e tecnologias de atendimento. Já *telemarketing* ou *televendas* é utilizado para o nicho de vendas e marketing, enquanto *contact center* é amplamente utilizado para denominar centros com vários tipos de comunicação, que além do telefone, suportam sistemas de mensagens, *chat* e *e-mail* (Ribeiro *et al.*, 2015).

A tecnologia pode desempenhar um papel crucial na estruturação e funcionamento de ambos os tipos de operações. Em *call centers* ativos, a tecnologia é fundamental para automatizar o processo de contato com o cliente. Sistemas de discagem preditiva, por exemplo, permitem realizar chamadas em massa, otimizando o tempo dos operadores e direcionando-os para potenciais clientes. Softwares de *Customer Relationship Management* (CRM) armazenam dados referente a informações dos clientes, que atua na personalização do atendimento e na realização de ofertas. Ferramentas de análise de dados e de automação de marketing também são amplamente utilizadas para segmentar clientes, criar campanhas segmentadas e otimizar as estratégias de comunicação (Ribeiro *et al.*, 2015).

Em *call centers* receptivos, a tecnologia pode garantir uma experiência mais fluida e eficiente ao cliente. Sistemas de espera virtual informam o tempo estimado de espera, evitando frustrações e proporcionando uma experiência mais transparente. Softwares de atendimento automatizado *Interative Voice Response* (IVR) permitem que o cliente acesse informações

básicas, como horários de funcionamento e Serviço de Atendimento ao Cidadão (SAC), sem a necessidade de conversar com agente de atendimento. A existência das plataformas de integração com os meios de comunicação (telefone, *chat* e *e-mail*) amplia as possibilidades de contato e oferece um atendimento eficiente (Ribeiro *et al.*, 2015).

Até a década de 1950, os *call centers* operavam de maneira precária e sem grandes avanços. No entanto, após a Segunda Guerra Mundial, especialmente nos Estados Unidos, onde os jornais eram amplamente difundidos, começou uma era de revolução no mercado. Através desses meios de comunicação, os anúncios permitiam que os leitores solicitassem produtos, serviços e suporte. A pioneira nesse sentido foi a empresa Ford, que montou uma equipe de aproximadamente 14 mil donas de casa. Elas realizavam ligações a partir de suas residências, alcançando milhões de contatos diários para entender o mercado alvo de automóveis da marca (Mancini, 2006).

O autor ainda destaca que nos Estados Unidos na década de 1970 aproximadamente 50% dos americanos que receberam contatos por telefone (vendas ou pesquisas) ouviam e se interessavam pelas propostas e ofertas. Na década de 1980 foi popularmente difundido o termo *telemarketing* nos EUA e outras regiões do mundo, principalmente no Brasil com a chegada das multinacionais americanas a partir da década de 1990.

O surgimento dos primeiros centros de atendimento no Brasil foi impulsionado pelas privatizações no setor de telecomunicações durante a segunda metade dos anos 1990. A priori, as empresas concentraram suas operações nas principais metrópoles do Sudeste, especialmente em São Paulo. Contudo, a partir do início de 2010, houve uma expansão geográfica do setor, outras regiões do país em busca de vantagens, como custos de mão de obra e benefícios fiscais, receberam diversas instalações de empresas do segmento. Esse movimento marcou o início das instalações de centros de atendimento no Nordeste (Mocelin; Silva, 2008).

Com o alto crescimento do setor no Brasil, houve a necessidade de surgimento de um órgão para estruturação do segmento no país, e com isso, em 1987 foi fundada a Associação Brasileira de Telesserviços (ABT), uma organização não governamental com propósito de representar e regulamentar o setor de *contact center* no Brasil. Desde 2005 é representada através da Programa Brasileiro de Autorregulamentação (Probare), que representa todos os serviços relacionados a *contact center* em território nacional. É uma iniciativa que engloba, além da ABT, outras duas entidades representativas do mercado de atendimento no país, a Associação Brasileira de Marketing Direto (BEMD) e Associação Brasileira das Relações Empresa Cliente(ABRAREC) (Probare, 2024).

Segundo Probare (2024), a iniciativa é gerenciada pela união do *board* executivo das três entidades, com a função de sustentar quatro vias:

- Código de ética: estabelece diretrizes de autorregulamentação para o setor de relacionamento com clientes, em conformidade com a legislação vigente. Desde 21 de novembro de 2019, é seguido por empresas do setor, visando melhorar o atendimento e alinhado às necessidades do mercado. A responsabilidade pela aplicação das normas cabe às empresas e suas centrais de relacionamento, abrangendo serviços como atendimento ao consumidor e suporte técnico;
- Ouvidoria: trata formalmente das violações ao Código de Ética;
- Selo de Ética: certifica que a organização está em conformidade com o Código de Ética. Pode ser obtido por qualquer central de atendimento, seja ela própria ou prestadora de serviços.
- Norma de Maturidade de Gestão: visa promover o aprimoramento contínuo das empresas do setor de relacionamento com clientes, aumentando a qualidade através de procedimentos e boas práticas.

Segundo Clientesa apud Freshworks (2023):

> As empresas brasileiras lideram mundialmente a relação de problemas solucionados logo no primeiro contato do consumidor, com média de 83% dos tickets resolvidos pela empresa especializada em softwares para CX, produzido a partir de 118 milhões de tickets de suporte de mais de 7.400 empresas de 106 países diferentes. O levantamento coloca o Brasil como país referência em assertividade no atendimento ao consumidor, já que a média global de resolução no primeiro contato é de apenas 72%.

Os *call centers* brasileiros têm tempo médio de resolução de 17,93 horas, posicionando o país como o segundo mais rápido do mundo no endereçamento de tickets. A Índia ocupa o primeiro lugar com um tempo médio de resolução de 17,12 horas. Por outro lado, a Itália está mal classificada, com um tempo médio de resolução de 35,23 horas para resolver problemas de consumo, o mais alto entre os países estudados. O tempo médio global de resolução é de 25,6 horas (Clientesa, 2023).

O principal desafio das centrais de atendimento é proporcionar um atendimento rápido e eficaz para seus clientes, visto que, a interação ágil proporciona maior produtividade da operação e custos menores. A previsão de demanda é uma ferramenta poderosa para o aumento da qualidade do atendimento e a redução de custos, pois interfere diretamente na alocação de recursos humanos e tecnológicos (Mandel; Feigin; Noiman, 2009).

1.1 JUSTIFICATIVA E MOTIVAÇÃO

A escolha do setor de *call center* foi motivada principalmente pela minha experiência de aproximadamente dois anos como analista de dados em operações de *call center* e atendimento. Esse histórico me proporcionou um conhecimento sólido sobre o funcionamento, os principais indicadores e os desafios característicos do setor, o que é fundamental para compreender e ajustar modelos de previsão de demandas em séries temporais com maior precisão.

Além disso, a decisão de segmentar para *call centers* receptivos foi influenciada pela alta variabilidade no volume de ligações recebidas, resultado de diversos fatores que impactam diretamente o dimensionamento da equipe e a manutenção dos níveis de serviço. A complexidade pode ser causada pela presença de múltiplos níveis de exigência no atendimento e à imprevisibilidade do volume de ligações, em contraste inverso ao de ambientes de *call centers* ativos, onde as listas de chamadas são pré-definidas (Koole; Siqiao; Wang, 2019).

Com base nesse cenário, o modelo em um ambiente mais desafiador pode alcançar maior potencial para ser replicado em outros contextos. Essa maior complexidade nas operações de *call center* tem fomentado diversos estudos que exploram desde modelos estatísticos clássicos até abordagens mais avançadas, como modelos de aprendizado de máquina e redes neurais.

Com as diversas abordagens possíveis na aplicação para previsão de volume de chamadas em *call center*, torna-se importante realizar um estudo evidenciando as metodologias, bem como, o detalhamento da teoria, funcionamento, vantagens e desvantagem de cada uma delas.

1.2 OBJETIVOS

No contexto evidenciado, o objetivo principal do trabalho é realizar uma pesquisa aplicada, exploratória e explicativa, através da verificação de modelos estatísticos clássicos e aprendizagem de máquina já estudados, e após, escolher e desenvolver um modelo composto por uma modelagem clássica e uma de aprendizagem de máquina para previsão de volume de chamadas em dados reais de um *call center*.

1.2.1 Objetivos Específicos

Para alcançar o objetivo central do trabalho, propõe-se realizar os seguintes tópicos:

- Entender o processo de previsão de chamadas de um *call center*, bem como

modelos comumente utilizados;

- Obtenção base de dados de um *call center*;
- Tratamento e limpeza dos dados coletados;
- Análise exploratória dos dados;
- Priorização de dois modelos, um clássico e um de aprendizagem de máquina mais frequentemente utilizado;
- Desenvolvimento das abordagens considerando as informações levantadas e as devidas comparações entre elas.

1.2.2 Estrutura do Trabalho

Este trabalho está estruturado em cinco capítulos. O Capítulo 1 oferece uma introdução abrangente sobre o tema, explorando uma visão geral dos *call centers* e a importância da previsão de chamadas nesse contexto. Também aborda os principais marcos históricos do setor, discutindo problemas e desafios enfrentados, além de estabelecer os objetivos que serão desenvolvidos ao longo do estudo.

No Capítulo 2, são apresentados os fundamentos teóricos que sustentam o desenvolvimento deste trabalho. Esse capítulo detalha o funcionamento de um *call center* e os desafios atuais enfrentados pelo setor. Além disso, explora os principais métodos clássicos e de aprendizado de máquina empregados na previsão de chamadas. Também é realizada uma revisão abrangente de estudos e pesquisas anteriores sobre previsão de chamadas em *call centers*, destacando contribuições relevantes e identificando lacunas que este estudo busca abordar.

No Capítulo 3 apresenta detalhadamente a metodologia utilizada para a execução da parte prática deste trabalho. Nele, são definidas as etapas e o fluxo adotado para as modelagens, além de descrever as ferramentas empregadas para o desenvolvimento das previsões.

No Capítulo 4, são expostos os resultados obtidos durante o processo de previsão, abrangendo desde as etapas iniciais de limpeza e análise exploratória dos dados até as previsões realizadas com as modelagens ARIMA e *Random Forest*. Esse capítulo aprofunda as análises, ponderações e resultados obtidos, detalhando o desempenho das modelagens aplicadas.

Por fim, o Capítulo 5 traz as conclusões do trabalho, sintetizando os principais achados e sugerindo possíveis melhorias e direções para estudos futuros sobre o tema, visando a continuidade e aprimoramento da previsão de chamadas em *call centers*.

2 FUNDAMENTAÇÃO TEÓRICA e REVISÃO DA LITERATURA

2.1 CALL CENTERS

O *call center* é considerado um ambiente de fluxo de trabalho contínuo, caracterizado pela entrada de ligações realizadas por consumidores que inicialmente entram em uma fila de espera, até que uma agente de atendimento com as qualificações necessárias recepcione a chamada. Esse ambiente é caracterizado com elevada taxa de imprevisibilidade quanto a chegadas de novas ligações, fazendo com os colaboradores tenham picos de estresses recorrentes (Cleveland; Mayben, 1997).

Segundo Cleveland e Mayben (1997), as centrais de atendimento estão avançando em complexidade ao decorrer do tempo, através do surgimento de customização dos atendimentos devido a públicos cada vez com mais pluralidades. Além das exigências de qualificações abrangentes por parte dos funcionários, houve a necessidade de existir tecnologias capazes de automatizar rotinas e gerenciar o direcionamento das ligações e interações entre empresa e clientes.

Para Bergevin *et al.* (2010), um *call center* pode ser definido a partir de duas vertentes. A do consumidor, que entende o *call center* como uma central com várias pessoas dispostas em filas com um telefone ao seu lado (conhecidos com atendentes, associados, consultores ou representantes), que tem a função de recepcionar ligações para oferta de algum tipo de serviço, seja de vendas ou suporte. Do outro lado, há a definição das companhias, que enxergam o *call center* como um centro estratégico para relacionamento com clientes, que envolvem fontes de receitas, custos, campanhas publicitárias, testes de consumo e *marketing* paralelo.

2.1.1 Call Center Inbound

O *call center* receptivo funciona com base na chegada de chamadas dos clientes, que geralmente são aleatórias e direcionadas para uma linha telefônica gratuita. O processo envolve encaminhar os clientes para os agentes disponíveis ou colocá-los em uma fila se nenhum agente estiver livre. Posteriormente, o cliente interage com um agente por um período até que a chamada seja encerrada, embora possam surgir complicações. Os clientes podem enfrentar obstáculos, como encontrar um sinal de ocupado ou optar por se desconectar enquanto estão

em espera (Mehrotra; Grossman; Samuelson, 2011).

Após a interação inicial, os clientes podem ser transferidos para outro agente ou ficam em uma fila para obter suporte adicional. Em alguns casos, o problema do cliente pode permanecer sem solução, levando a uma futura ligação de acompanhamento.

Como o tempo de espera variam entre os clientes, as métricas são influenciadas pela distribuição deste indicador. As principais métricas usadas são a *Average Speed of Answer* (ASA) e *Service Level* (SL), indicando o tempo médio de espera do cliente e a porcentagem de chamadas atendidas em um período especificado, respectivamente. A atenção recente de pesquisadores e gerentes mudou para a renegação ou abandono do cliente, com a *Call Answer Rate* (CAR) emergindo como uma métrica crítica (Cleveland; Mayben, 1997).

Durante o processo acontecem de clientes insatisfeitos abandonam a fila, que afetam negativamente na relação empresa e cliente. As métricas relacionadas ao tempo de espera e ao abandono refletem a experiência de pré-atendimento do cliente, em que, cada vez mais cresce o interesse em avaliam a qualidade e experiência do cliente durante a prestação de serviços, incluindo agentes individuais e grupos de agentes (Mehrotra; Grossman; Samuelson, 2011).

Em *call center inbound*, a taxa de chagada de ligações podem variar de acordo com vários fatores, de acordo com Cleveland e Mayben (1997) há dois impactos principais: aleatórios e picos de sazonalidade. O efeito aleatório pode ser observado em análises intradiárias com a presença de variações atípicas e não padronizadas no contexto histórico. De outro modo, a sazonalidade tem grande influência e geralmente pode ser prevista, pois é relacionado a eventos pré-estabelecidos ou tendências de mercado.

2.1.2 Sistemas Outbound

Para Bergevin *et al.* (2010), um *call center outbound* é denominado uma operação ativa, pois o contato inicial com clientes é de iniciativa da própria central de atendimento, é utilizado uma lista pré-selecionada de contatos, para tratar de assuntos de cobranças, pesquisas de satisfação ou realização de campanhas de *marketing*. O custo embarcado em sistemas de *outbound* geralmente são superiores ao *inbound*, dado que em diversas situações há a realização de vários contatos com o mesmo cliente, e a necessidade de acompanhamento de toda essa jornada por parte do prestador de serviços ou produto.

Há um alto volume de ligações realizadas por dia e, também de rejeições destas por parte dos clientes, o processo na grande maioria dos *call center* desta modalidade é realizada

por *software* especializados, que geram as listas personalizadas considerando critérios definidos pelas áreas de negócio e fazem todo o redirecionamento de cada ligação atendida para os agentes de atendimento (Mehrotra; Grossman; Samuelson, 2011).

2.1.3 Gerenciamento e funcionamento

Em operações de *call center* receptivo o custo referente a recurso humanos ainda é elevado, mesmo com diversos avanços tecnológicos. Segundo Koole, Gans e Mandelbaum (2003), existem duas razões interligadas que sustentam a ênfase no gerenciamento de capacidade. Na maioria dos *call centers*, os custos associados à capacidade, especialmente aqueles relacionados aos recursos humanos, constituem 60% a 70% do total das despesas operacionais. Além disso, uma parte significativa dos estudos existentes se concentrou no tema do gerenciamento de capacidade. Essa tendência provavelmente decorre da prioridade histórica dada à pesquisa de gerenciamento de operações, bem como do reconhecimento dos pesquisadores da importância financeira associada às despesas de capacidade.

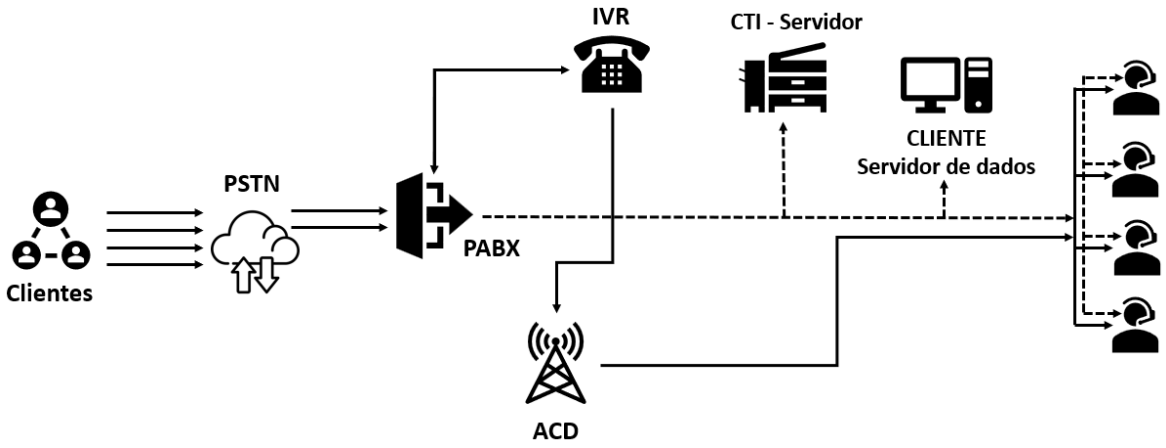
A complexidade do gerenciamento vem aumentando com o surgimento de necessidades de uma compreensão mais profunda da influência dos fatores humanos e da otimização de tecnologia, como redes e mecanismos projetados para roteamento baseado em habilidades dos agentes de atendimento. As questões relacionadas ao comportamento e à tecnologia estão intrinsecamente interconectadas (Koole; Siqiao; Wang, 2019).

Conforme já abordado, sabe-se que há duas ramificações importantes para *call center*, *inbound* e *outbound*. O foco do trabalho vai ser *call center inbound*, devido a aplicabilidade proposta e pela individualidade de características de cada um.

2.1.4 Distribuição de Chamadas

A distribuição de ligações pode variar de operação para operação em um *call center*, algumas características são cruciais para determinação do modelo, como dimensões, local de instalações, quantidade de tecnologias embarcadas e serviços oferecidos. A fim de simplificar, a Figura 1 é possível observar a estrutura base de como é o fluxo desde a efetuação da ligação por parte do cliente até o atendimento dos agentes do *call center*.

Figura 1 - Diagrama de entrada de chamadas em call center inbound



Fonte: Adaptado de Koole, Gans e Mandelbaum (2003).

Segundo Koole, Gans e Mandelbaum (2003), quando os clientes efetuam alguma ligação, o primeiro direcionamento é para a rede conhecida como *Public Switched Telephone Network* (PSTN). A infraestrutura da rede é baseada em componentes físicos, como fios de cobre, cabos de fibra óptica, centrais de comutação, redes celulares e satélites, essa rede tem o objetivo principal de conectar origem e destino de cada chamada. Duas tecnologias embarcadas são extremamente importantes para a conexão entre os dois pontos, o *Automatic Number Identification* (ANI), que serve para identificação automática do número e localização da origem, e o *Dialed Number Identification Service* (DNIS) que é o serviço de identificação do destinatário da ligação e roteamento correto das chamadas. Após as devidas identificações há a conexão entre a clientes o centro de atendimento.

Koole, Gans e Mandelbaum (op. cit) destaca que após realizado as identificações de origem e destino, as chamadas entram em uma rede privada, que pertence a empresa de atendimento, conhecida por *Private Automatic Branch Exchange* (PABX), ilustrada na Figura 1. Para que a conexão seja bem-sucedida, a PABX deve ser configurada com uma capacidade pré-estabelecida. Caso essa capacidade seja excedida, as pessoas que tentarem conectar-se à central receberão um sinal de ocupado.

De acordo com referido autor, a ligação inicialmente é encaminhada ao IVR, que é um servidor de respostas automáticas e programadas, são caracterizados por uma voz robótica que guia os clientes em diversas opções, caso este serviço não consiga atender as necessidades, caso este cliente deseje, há possibilidade de ser direcionado para um agente de atendimento. Para isso, a ligação é conectada do IVR para A*utomatic Call Distributor* (ACD), que tem a função de distribuir as ligações para os agentes de atendimento através de diversos critérios, como

idioma falado pelo discador, habilidades de cada agente, entre outros.

Koole, Gans e Mandelbaum (2003) ainda destaca que o funcionamento de todo o fluxo explicado acima depende do *Computer Telephone Integration* (CTI) e o servidor de dados da empresa. O CTI é responsável por coletar todas as informações em tempo real durante a ligação, abrangendo toda a jornada do cliente desde a primeira interação até o momento de conexão com os agentes. Ele está conectado ao servidor de banco de dados e aos sistemas da empresa, permitindo que o agente de atendimento tenha acesso a todas as informações dos clientes, incluindo os dados históricos.

2.1.5 Infraestrutura e Tecnologia

Sabe-se que o capital humano é crucial para o funcionamento de um *call center*, representando mais da metade do orçamento, contudo, a infraestrutura também desempenha um papel fundamental.

Segundo Pinedo, Seshadti e Shanthikumar (2000) o IVR, ACD e CTI Servidor são as principais tecnologias para bom funcionamento do fluxo de trabalho em um *call center*. O IVR é uma interface eletrônica baseada em uma arvore de opções que realiza a primeira interação entre cliente e *call center*, esses itens são previamente definidos de acordo com a operação e funcionam de maneira automatizada. As interações podem ser de duas maneiras, via digitação e/ou voz, dependendo do nível de tecnologias embarcadas. A utilização de IVR é vetor principal para diminuição de filas, entretanto, há situações que a interface não consegue suprir as necessidades de algumas pessoas, portanto, há uma opção de transferência para agentes de atendimento disponíveis no momento da ligação.

Conforme o autor citado acima, o ACD recebe as ligações advindas do IVR, que tem a função de distribuir as chamadas considerando todo o histórico da ligação. O redirecionamento é seletivo, considerando vários fatores, como por exemplo, geográficos e características das solicitações.

Para o funcionamento do IVR e ACD, é crucial a existência do CTI, sua origem deve-se a necessidade de integração entre computador e telefone. Conforme fluxo da Figura 1, o CTI tem a função de capturar as informações em tempo real durante o contato com clientes e: realizar o armazenamento desses dados, correlacionar com dados históricos e fornecer para os agentes de atendimento (Sharp, 2003).

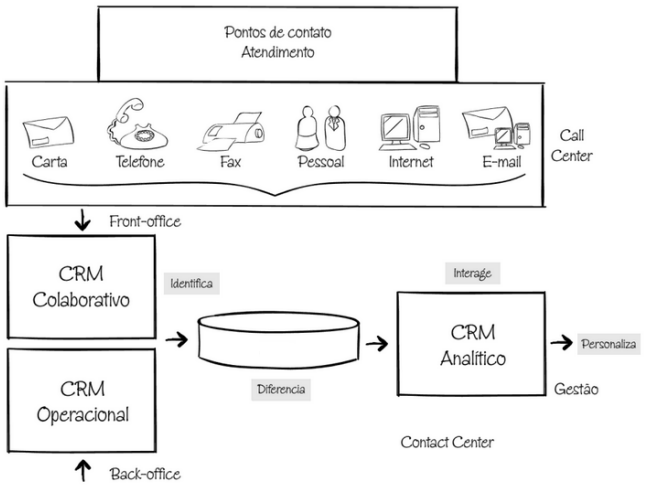
Nos primórdios eram utilizado um computador com configurações mais avançadas

chamado de servidor, este se comunicava com os demais computadores via rede interna da empresa. Com a modernização e avanços tecnológicos, atualmente há duas opções que são utilizadas, a contratação da infraestrutura de CTI em nuvem, ou seja, um servidor em nuvem que executa os serviços necessários, ou a contratação de plataformas que oferecem o serviço desde o contato inicial do cliente até o atendimento (Sharp, 2003).

Com o aumento das operações de atendimento, identificou-se a necessidade de tecnologias que integrasse as informações dos clientes, surgindo então, o CRM, um sistema de gerenciamento de clientes através da utilização de todas as informações do CTI, porém com o aditivo de geração de *insights* valiosos para os agentes de atendimento. Além da integração das informações, o CRM estreita as relações entre empresa e cliente, através da personalização dos atendimentos e velocidade para resolução de conflitos (Zenone, 2007).

Ainda segundo o autor citado, o CRM é uma poderosa ferramenta que além de integrar as diversas aplicações das empresas e fornecer análises descritivas dos clientes, é utilizada para geração de predições de comportamentos que ajudam a melhorar a qualidade de atendimento. Para isso, o CRM é dividido em três aplicações conforme Figura 2.

Figura 2 - Divisão de aplicabilidade do CRM



Fonte: Zenone (2007).

O CRM colaborativo é responsável por receber as informações dos meios de integração empresa cliente. O CRM operacional engloba as automações de fluxos de planejamento, organização das filas de atendimento e histórico de dados dos clientes, este proporciona a qualificação do atendimento e rapidez na resolução de solicitações. Já o CRM analítico é a fonte de inteligência de todo o processo, fornecendo *insights* a partir dos dados disponíveis e

disponibilizando em uma interface amigável para os consultores de atendimento (Zenone, 2007).

2.1.6 Métricas Utilizadas em Call Center Receptivo

Segundo Franzese *et al.* (2009), em um *call center* receptivo analisa-se as seguintes variáveis:

1. Quantidade de chamadas: ligações que conseguem ser completadas até o IVR;
2. Quantidade de chamadas finalizadas: ligações que são resolvidas pelo IVR ou recepcionadas pelos agentes de atendimento;
3. Média de pessoas em fila: média de clientes que ficam aguardando atendimento no ACD;
4. Tempo médio de espera: tempo médio em que uma pessoa aguarda para ser atendido por um agente de atendimento;
5. Taxa de abandono: porcentagem de chamadas que são abandonadas sem resolução no ACD;
6. Taxa de ocupação: porcentagem de agentes de atendimento ocupados na operação;
7. Taxa de chamadas com sinal ocupado: porcentagem de chamadas que não conseguem chegar até o IVR, devido a indisponibilidade de linha telefônica, isto ocorre quando a empresa possuí um número de linhas menor que a demanda recebida;
8. Custo por chamadas: custos operacionais fixos e variáveis dividido pela quantidade de chamadas recebidas, considerando um período específico.

De acordo com Koole, Gans e Mandelbaum (2003), em uma central de atendimento dimensionada corretamente, a taxa de ocupação varia entre 90% e 95%, nenhum cliente recebe sinal de ocupado, e mais da metade da demanda é atendida imediatamente, enquanto o restante espera por pouco tempo na fila. Além disso, a taxa de abandono seria baixa, variando entre 1% e 2%. No entanto, os próprios autores mencionam que poucos *call centers* possuem uma estrutura ideal.

2.2 DESAFIOS PARA PREVISÃO DE CHAMADAS EM CALL CENTERS

A previsão de chamadas em *call center* é muito importante para otimização dos recursos, visto que proporciona melhorias de processos e planejamento para alocação de pessoas na operação. Porém, possuem desafios significativos que podem comprometer a eficiência e o desempenho, abaixo encontra-se a listagem dos principais desafios.

2.2.1 Variabilidade e aleatoriedade

De modo geral os *call centers* sofrem com a influência de fatores aleatórios advindo da variabilidade na chegada e duração das chamadas. A eficiência de previsão pode ser afetada pela generalização da modelagem para períodos maiores. À medida que tentamos prever para períodos mais longos, como meses e trimestres, a variabilidade entre esses períodos e períodos mais curtos, como dias e semanas, torna-se significativa (Cleveland; Mayben, 1997).

Segundo o autor citado, essa diferença de granularidade pode introduzir ruído e incerteza nas previsões, uma vez que fatores que influenciam tendências de curto prazo podem não ser os mesmos que afetam tendências de longo prazo. Além disso, modelos que se ajustam satisfatoriamente aos dados diários ou semanais podem não capturar adequadamente padrões mensais ou trimestrais, levando a previsões menos precisas.

Devido à natureza aleatória e imprevisível na chegada de ligações na operação, prever a demanda em tempo real é um desafio significativo, também. Embora previsões diárias e semanais possam ser satisfatórias, há uma grande dificuldade em determinar a demanda intradiária em intervalos de minutos (Koole; Gans; Mandelbaum, 2003);

Segundo Cleveland e Mayben (1997) as previsões em tempo real são prejudicadas principalmente pela variação no tempo das chamadas, que sofre influência de fatores difíceis de prever nos modelos tradicionais. Isso incluem variações no comportamento dos clientes, eventos inesperados e flutuações no volume de chamadas, podendo tornar a previsão precisa um objetivo complexo e desafiador.

2.2.2 Fatores externos

A influência de fatores externos no desempenho de um *call center* é um aspecto crítico a ser considerado na gestão dessas operações. Fatores como mudanças econômicas, eventos

sazonais e crises imprevistas podem impactar significativamente o volume e a natureza das chamadas recebidas. Por exemplo, em períodos de recessão econômica, pode haver um aumento nas chamadas relacionadas a cobranças e renegociações de dívidas. Eventos sazonais, como férias ou promoções específicas, também podem gerar picos de demanda que exigem uma gestão rápida e eficiente dos recursos disponíveis no *call center* (Vaughn, 2014).

Segundo Vaughn (op. Cit), além das variáveis econômicas e sazonais, fatores externos como desastres naturais, pandemias e mudanças nas regulamentações governamentais podem alterar drasticamente a dinâmica de um *call center*. A pandemia de COVID-19, por exemplo, levou a um aumento súbito no volume de chamadas em muitos setores, especialmente na área de saúde e serviços essenciais. Isso exigiu uma rápida adaptação por parte dos *call centers* para gerenciar o aumento de demanda, muitas vezes com recursos limitados devido às restrições impostas pela própria pandemia.

O autor reforça que fatores tecnológicos e sociais também desempenham um papel importante na influência externa sobre os *call centers*. A evolução das tecnologias de comunicação e o aumento do uso de canais digitais, como *chatbots* e redes sociais, modificam a forma como os clientes interagem com as empresas. Ao mesmo tempo, mudanças nas expectativas dos consumidores em relação ao tempo de resposta e à qualidade do atendimento exigem que os *call centers* se adaptem constantemente. Essas influências externas, portanto, exigem uma abordagem estratégica e proativa para garantir que os *call centers* possam atender de maneira eficaz às necessidades dos clientes, mantendo a eficiência operacional.

2.2.3 Oportunidades para Aprimorar a Previsão

Segundo Anton (1997), o gerenciamento da qualidade de serviço é um aspecto fundamental para otimizar o processo de previsão de demanda em operações de atendimento. Em um ambiente de *call center*, a precisão na previsão de demanda não só garante que os recursos sejam alocados de maneira eficiente, mas também assegura que a experiência do cliente seja consistentemente positiva. Ao entender as necessidades futuras e ajustar os recursos de acordo, a operação consegue minimizar tempos de espera, reduzir a sobrecarga dos agentes e melhorar a satisfação geral dos clientes. Assim, a previsibilidade torna-se um elemento chave na construção de um atendimento de alta qualidade.

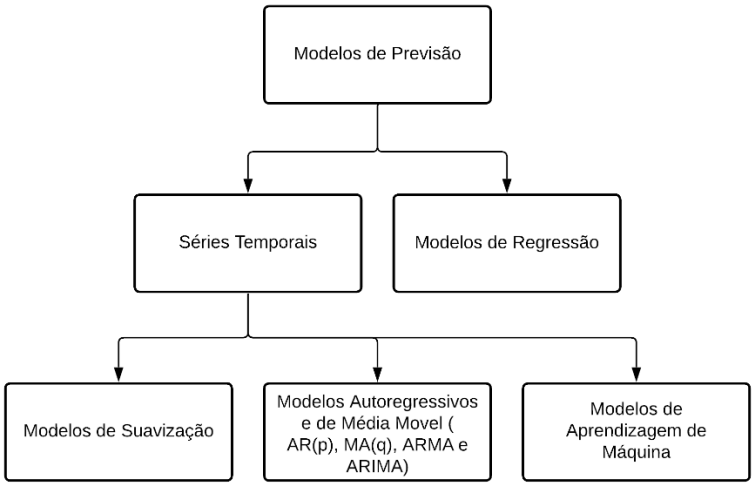
Para que essa previsibilidade seja eficaz, é essencial que o processo de previsão de demanda esteja bem definido e alinhado com as demais áreas, como treinamento, sistemas de

informação, recursos humanos e os times operacionais. Quando esses setores trabalham de forma sinérgica, é possível criar um fluxo de informação constante e preciso, que permite ajustes rápidos e precisos nas operações. A integração dessas áreas contribui para a identificação de padrões de demanda, possibilitando a antecipação de possíveis picos aleatórios e garante que a equipe esteja preparada para lidar com eles de maneira eficiente (Anton, 1997).

2.3 PREVISÃO DE VOLUME DE CHAMADAS

Atualmente existem diversas abordagens para previsão de chamadas em *call center*, desde modelos de regressões simples até utilização de modelos estatísticos e aprendizagem de máquina para séries temporais, uma visão ampla pode ser visualizada a partir da Figura 3.

Figura 3 - Modelos de Previsão



Fonte: Elaborado pelo autor (2024).

2.3.1 Métodos Estatísticos

Antes de nos aprofundarmos nos modelos estatísticos, é essencial definir alguns conceitos básicos que servirão como base para a compreensão mais avançada. Esses conceitos incluem a definição do que é estatística, a importância dos dados e como eles são coletados, organizados e analisados. Somente após estabelecermos esses fundamentos, poderemos explorar os modelos estatísticos de maneira mais eficaz.

A estatística pode ser entendida como o estudo do estado de um evento ou fenômeno,

utilizando-se de experimentos, coleta e organização de dados, seguidos pelo resumo, apresentação e análise interpretativa dessas informações. Com base nesse conjunto de dados, são feitas as devidas considerações e inferências. É considerada uma ciência analítica, pois emprega técnicas de compilação de dados, gráficos e modelos matemáticos para descrever e explicar padrões e tendências observados nas bases estudadas. Além disso, a estatística é fundamental para a tomada de decisões em diversas áreas, como economia, saúde e ciências sociais, permitindo que se façam previsões e avaliações de riscos com base em dados concretos (Triola, 2017).

Os dados podem ser coletados a partir de duas fontes distintas: estudos observacionais e experimentos. Em estudos observacionais, os dados são obtidos mediante a identificação e análise de eventos ou fenômenos, sem que haja qualquer intervenção direta sobre as variáveis em estudo. Este tipo de abordagem permite que se capturem informações sobre o comportamento natural dos sujeitos ou sistemas observados. Por outro lado, os experimentos são delineados com o objetivo de avaliar os impactos de intervenções específicas sobre as unidades de análise, frequentemente denominadas unidades experimentais. Nesse contexto, o pesquisador manipula uma ou mais variáveis independentes, enquanto controla outras possíveis fontes de variação, a fim de observar os efeitos causais diretos sobre as variáveis dependentes (Triola, 2017).

A coleta de dados é realizada principalmente através de técnicas de amostragem, que envolvem a seleção de uma parte representativa da população para possibilitar inferências sobre o conjunto total de indivíduos em estudo. A população, nesse contexto, refere-se ao universo completo dos sujeitos ou elementos a serem investigados. Entre as diversas abordagens de amostragem, a amostragem aleatória simples é a mais comumente utilizada e considerada robusta. Nessa técnica, os indivíduos são selecionados de forma completamente aleatória, garantindo que cada elemento tenha a mesma chance de ser escolhido. Esse processo minimiza o viés, que é a inclinação ou distorção nos resultados devido a uma seleção inadequada ou não representativa da amostra, e aumenta a representatividade da amostra, assegurando a validade e a confiabilidade das inferências estatísticas que se seguem (Bolfarine, 2005).

Para a análise dos dados, antes de se aplicar modelos estatísticos mais complexos, pode-se realizar a estatística descritiva. Essa etapa inicial inclui a utilização de análises gráficas, como distribuições de frequências, histogramas, diagrama de caixa (gráfico *boxplot*) e gráficos de correlação, além de comparações de medidas como médias, mediana, quartis, variações e covariância. Essas técnicas permitem um entendimento preliminar dos padrões e tendências

presentes nos dados, facilitando a identificação de características importantes e potenciais relações entre as variáveis antes de se avançar para análises inferenciais e para uma modelagem estatística mais avançada (Oore; Notz; Fligner, 2023).

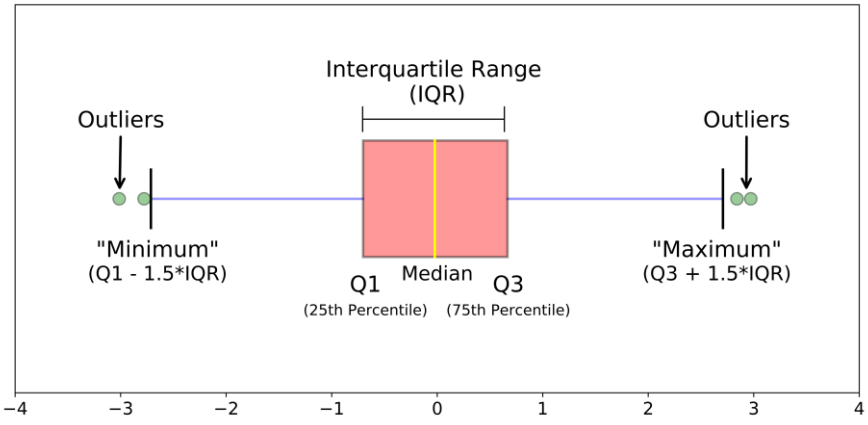
A mediana é o valor que divide um conjunto de dados ordenado ao meio, correspondendo aos 50% centrais. Esse ponto também é conhecido como segundo quartil (Q2). Além de Q2, temos o primeiro quartil (Q1), que marca os 25% inferiores dos dados, e o terceiro quartil (Q3), que representa os 75% inferiores, dividindo os dados em quatro partes iguais (Morettin; Bussab, 2017).

Ainda segundo o autor, as métricas citada acima é utilizada para elaborar o gráfico *boxplot*, que é essencial para a visualização do comportamento dos conjuntos de dados e a identificação de outliers, valores atípicos que se destacam dos demais. Esses outliers geralmente representam erros ou situações específicas que não refletem o comportamento típico do conjunto de dados. Para identificação, são considerados os valores acima de um limite superior e abaixo de um limite inferior, representado na equação (1) e (2), onde IQR representa a diferença entre Q3 e Q1 e ilustrado pela Figura 4.

$$\text{Limite Inferior} = \text{Q1} - (1{,}5 \times \text{IQR}) \tag{1}$$

$$\text{Limite Superior} = \text{Q3} + (1{,}5 \times \text{IQR}) \tag{2}$$

Figura 4 - Gráfico *boxplot*



Fonte: Labone Consultoria e Treinamentos (2024).

A média corresponde a medida de centro através do somatório de um conjunto de valores dividido pela quantidade de total destes valores. Conforme demostrado na equação (3) comumente a média é representado por $\bar{x}$ para conjunto amostral e pela letra grega μ para conjunto populacional, e n equivale ao tamanho da amostra ou população (Triola, 2017).

$$média = \mu = \frac{\sum x}{n} \quad (3)$$

A variância é uma medida igual ao quadrado do desvio padrão, sendo representado por s^2 para conjunto amostral e σ^2 para uma população. O desvio padrão é definido pelas equações (4) e (5), e mede quantitativamente a defasagem entre um determinado valor com referência da média. (Triola, 2017)

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \quad (4)$$

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}} \quad (5)$$

Com s e σ é a representação do desvio padrão amostral e populacional, respectivamente. O valor de x representa os valores representantes de uma amostra ou população. E n e N o tamanho dos conjuntos amostrais ou populacionais.

Segundo Morettin e Bussab (2017) a covariância tem o objetivo de medir a direção do relacionamento linear entre duas variáveis (X, Y) que pode corresponder a qualquer valor real, fazendo com que varie de acordo com a escala das variáveis utilizadas. Matematicamente, a covariância é definida como:

$$\text{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \quad (6)$$

Onde $\text{Cov}(X, Y)$ é o valor calculado da covariância entre X e Y, X_i e Y_i valores amostrais de uma amostra de tamanho n, e $\bar{X}$e $\bar{Y}$ são as médias respectivas de X e Y.

Ainda segundo o autor citado, há três interpretação para covariância, elas são:

- Covariância positiva: quando a covariância é maior que zero, significa que as duas variáveis apresentam uma tendência de aumentar ou diminuir em conjunto. Em outras palavras, se uma variável experimenta um aumento, a outra variável também tende a experimentar um aumento.
- Covariância negativa: quando a covariância é menor que zero, isso significa que as

variáveis demonstram uma propensão a flutuar em direções opostas. Especificamente, quando uma variável experimenta um aumento, a outra variável tende a sofrer uma diminuição.

- Covariância igual a zero: indica que não existe associação linear entre as duas variáveis.

No contexto de previsão de demanda, especialmente em *call centers*, a estatística oferece dois campos principais que podem ser utilizados para a consolidação de previsões: a Regressão Linear (simples ou múltipla) e os métodos de Séries Temporais. A Regressão Linear é útil para identificar e modelar a relação entre a demanda e uma ou mais variáveis independentes, enquanto os métodos de Séries Temporais analisam os dados ao longo do tempo, capturando padrões e tendências sazonais ou cíclicas que são essenciais para prever a demanda futura com maior precisão.

2.3.1.1 Regressão

A regressão é um método estatístico que examina a relação entre duas ou mais variáveis, buscando aproximá-las uma função linear ou não linear. Em vez de uma relação determinística, a regressão linear se concentra em uma modelagem probabilísticos. Isso significa que, embora a equação da reta descreva um padrão geral, não determina completamente o valor de uma variável com base na outra sem a consideração de uma incerteza atrelada (CASTANHEIRA, 2013).

Segundo Triola (2017), existem dois tipos principais de regressão linear: a regressão linear simples e a regressão linear múltipla. A regressão linear simples utiliza dados amostrais emparelhados para identificar a existência de uma relação linear entre duas variáveis. A partir desses dados, é possível determinar uma reta de regressão, também conhecida como a reta de melhor ajuste dos dados observados, conforme equação (7).

$$\hat{y} = b_0 + b_1 x \tag{7}$$

Onde x é a variável independente ou preditora, b_0 valor de intercepção no eixo y, b_1 valor da inclinação da reta e $\hat{y}$ é a variável dependente ou resposta amostral.

A regressão linear múltipla é uma técnica estatística que estuda a relação entre uma variável dependente (variável resposta) e duas ou mais variáveis independentes (variáveis preditoras), conforme equação (8).

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k \quad (8)$$

Onde x_1, x_2, ..., x_k são as variáveis preditoras, b_0 valor de intercepção no eixo y, e b_1, b_2, ..., b_k são valores da inclinação da reta para cada variável preditora e $\hat{y}$ é a variável resposta.

Ainda segundo o autor citado, a regressão não linear é utilizada para modelar relações entre variáveis que não podem ser adequadamente representadas por uma linha reta. Ao contrário da regressão linear, a regressão não linear utiliza funções mais complexas (como polinômios, exponenciais ou logarítmicas) para capturar padrões nos dados.

2.3.1.1.1 Métricas de Avaliação Regressão

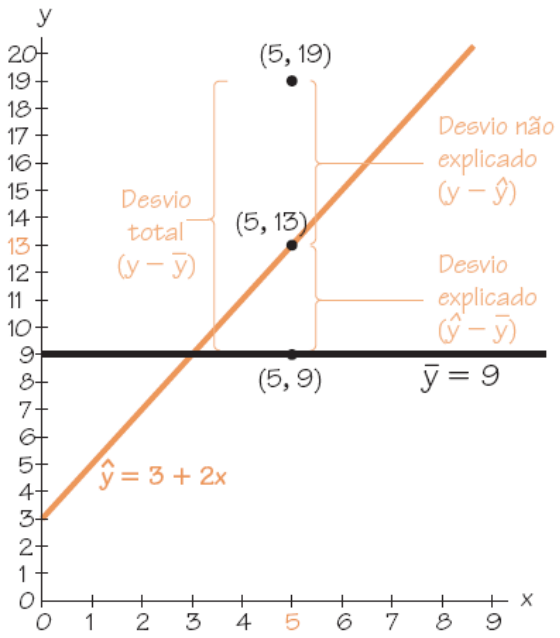
Para avaliação dos modelos de regressão linear as principais métricas utilizadas são: Raiz do Erro Quadrático Médio (*Root Mean Squared Error* RMSE) e Coeficiente de Determinação (R^2) e R^2 ajustado.

O RMSE ou Erro padrão da regressão é a medição da magnitude média do erro, na mesma unidade da variável dependente segundo a equação (9) (Oore; Notz; Fligner, 2023).

$$\text{RMSE} = \sqrt{\frac{1}{n - gl}\sum_{i=1}^{i}(y_i - \hat{y})^2} \quad (9)$$

Onde a expressão $(y_i - \hat{y})^2$ representa a soma dos quadrados dos resíduos, que é a diferença entre as observações y_i em relação a reta ajustada $\hat{y}$ representado pelo item: Desvio não explicado na Figura 5. O n refere-se ao número de observações, enquanto gl representa os graus de liberdade. A cada parâmetro estimado, como a intercepção no eixo y e a inclinação da reta, perde-se um grau de liberdade. Em uma regressão simples, conforme a equação (7), o gl equivale a 2. Quanto menor o RMSE, menor a diferença entre os valores preditos e os reais, sendo assim, é importante avaliar o contexto e as necessidades da predição para determinar se o modelo é adequado ou não.

Figura 5 - Representação para calcular o resíduo $y_i - \hat{y}$



Fonte: Triola (2017).

O Coeficiente de Determinação (R^2) é uma medida estatística que indica a proporção da variabilidade na variável dependente que é explicada pelas variáveis independentes em um modelo de regressão. Em outras palavras, R^2 quantifica o quanto do total das variações dos dados observados pode ser previsto pelo modelo e é definido a partir da equação (10) (Sicsú; Samartini; Barth, 2023).

$$R^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(\hat{y} - \bar{y})^2 + \sum_{i=1}^{i}(y_i - \hat{y})^2} \qquad (10)$$

O numerador é a soma do desvio explicado, que é a distância entre o valor predito $\hat{y}$ e a reta horizontal que passa pela média amostral $\bar{y}$, e o denominador, é soma entre desvio explicado e desvio não explicado, sendo o último equivalente ao resíduo abordado na equação (9).

Segundo Triola (2017), é possível calcular o Coeficiente de Determinação Ajustado (R^2 ajustado), que faz com que o R^2 convencional se ajuste ao número de variáveis preditoras no modelo. Geralmente, à medida que variáveis preditoras são adicionadas ao modelo, o R^2 tende a aumentar. No entanto, isso não significa necessariamente que todas essas variáveis contribuem de forma significativa para o ajuste ideal da reta de regressão. Para compensar esse

efeito e fornecer uma avaliação mais precisa da qualidade do modelo, utiliza-se esta métrica. O R^2 é utilizado considerando o número de variáveis preditoras no modelo (k) e o tamanho amostral (n), sendo definido pela equação (11).

$$R^2\ ajustado = 1 - \frac{(n-1)}{[n-(k+1)]}(1-R^2) \tag{11}$$

Outras duas métricas utilizadas é o Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error* MAPE) e (*Mean Absolute Error* MAE). O MAPE tem a função de avaliar a capacidade preditiva do modelo a partir do erro relativo, apresentando a magnitude do erro em formato percentual, em que valores baixos indicam menor discrepância entre valores reais e previstos, é calculado dividindo-se o resíduo (y_i - $\hat{y}_i$) pelo valor observado, onde y representa o valor observado e ŷ o valor previsto, e n o total de observações da amostra. O MAE é uma métrica mais intuitiva se comparado ao RMSE e não sofre com a presença de *outiliers* e mede a média absoluta dos resíduos. Ambas as equações são expressas abaixo (Sicsú; Samartini; Barth, 2023).

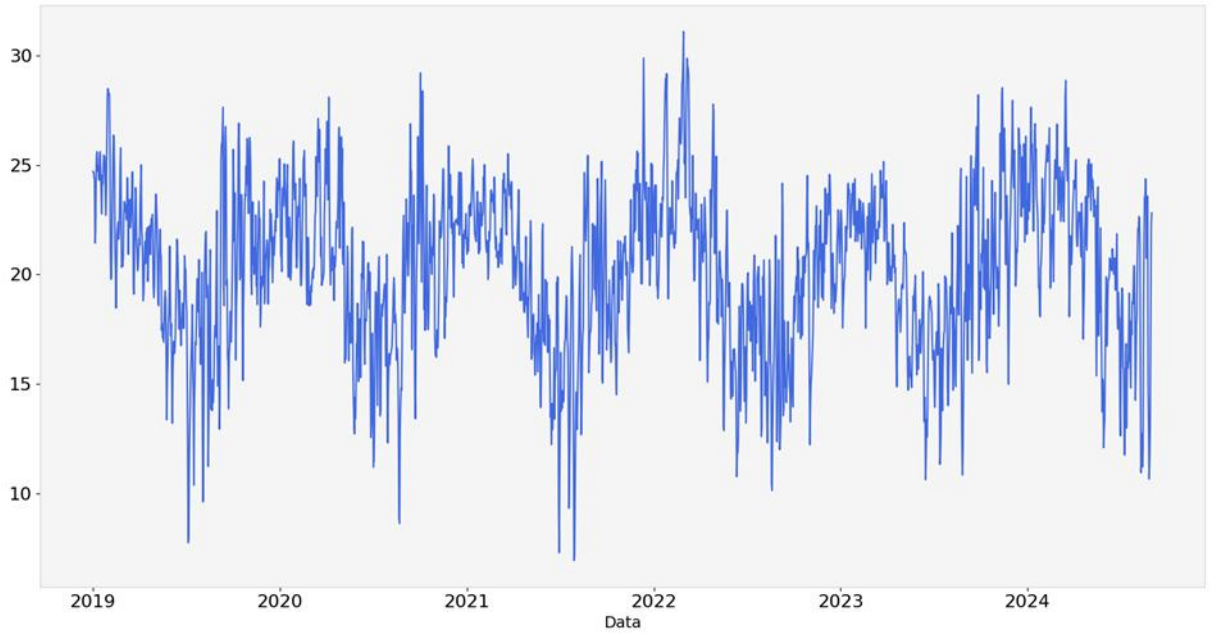
$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{12}$$

$$\text{MAE} = \sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{n} \tag{13}$$

2.3.1.2 Séries Temporais

Uma série temporal consiste na observação de um fenômeno ao longo de um período definido, podendo ser discreta ou contínua. Em séries temporais discretas, as observações são realizadas sequencialmente em tempo específicos, que variam de nanosegundos a séculos, como de dados de temperatura diária, conforme a Figura 6, que comtemplam temperaturas médias coletadas em Celsius em Itapeva/SP entre janeiro de 2019 e agosto de 2024.

Figura 6 - Temperatura Média Diária (°C) - Itapeva/SP



Fonte: Elaborado pelo autor (2024).

Em séries temporais contínuas, os dados variam de forma ininterrupta ao longo do tempo, sem uma escala temporal definida, como ocorre, por exemplo, em registros de encefalogramas de pacientes durante crises epiléticas. No entanto, é importante destacar que, ao coletar dados de uma série contínua, ocorre automaticamente uma discretização, pois é necessário introduzir uma escala temporal para sumarizar as informações. A análise de séries temporais pode ser univariada ou multivariada. Na modelagem univariada, utiliza-se apenas uma variável de interesse e a unidade temporal para fazer previsões. Já na modelagem multivariada, outras variáveis adicionais podem ser incorporadas para ajustar e melhorar a previsão, levando em consideração suas interações com a variável principal ao longo do tempo (Morettin; Toloi, 2018).

O mesmo autor destaca que, as séries temporais são caracterizadas como processos estocásticos, ou seja, processos que evoluem ao longo do tempo influenciados por elementos aleatórios. Essa natureza estocástica se manifesta na forma como cada observação Z(t) de uma série temporal é influenciada por variáveis aleatórias, tornando as previsões para o futuro probabilísticas e não determinísticas.

De maneira geral, o processo estocástico de uma série temporal é definido por um conjunto de observações $Z_{(t)}$ ao longo do tempo, onde cada observação é uma variável aleatória distribuída em uma sequência interligada por uma estrutura dependente do tempo.

As séries temporais podem apresentar observações irregulares, ou seja, os dados podem não seguir uma sequência regular no tempo. Isso ocorre quando a variável de interesse é registrada em certos intervalos de tempo, mas em outros não. Além disso, uma série temporal pode ser classificada como estacionária ou não estacionária. Séries temporais estacionárias são caracterizadas por uma média (μ) e variância (σ^2) constantes ao longo do tempo, e sua covariância ($\mathrm{Cov}\ (X_t, X_{t+h})$) dependendo apenas das diferenças temporais entre os valores, com X_t representando um ponto específico no tempo t comparado com outro ponto com defasagem de tempo h (Morettin; Bussab, 2017).

2.3.1.2.1 Componentes de Séries Temporais e Transformações

Hyndman e Athanasopoulos (2021) define uma série temporal com a presença das componentes de tendência, sazonalidade e um termo aleatório (ruído), conforme decomposição via Figura 7. São representadas por dois modelos principais, aditivo e multiplicativo, definidos pelas equações (14) e (15), respectivamente. A principal diferença entre os modelos é que o modelo multiplicativo é mais adequado quando a componente sazonal é influenciada pela tendência.

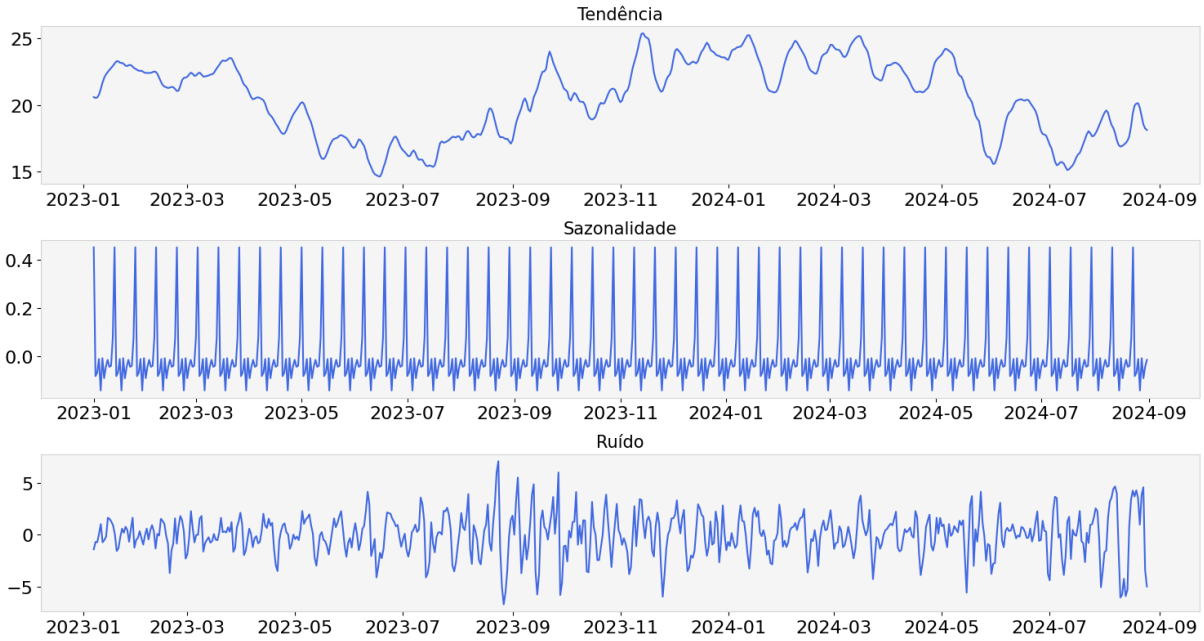
$$Y_t = T_t + S_t + R_t, t = 1, 2, \dots, n \quad (14)$$

$$Y_t = T_t \times S_t \times R_t, t = 1, 2, \dots, n \quad (15)$$

Onde:

- Y_t é o valor da série no tempo t;
- T_t é a tendência no tempo t;
- S_té o componente sazonal no tempo t;
- R_té o ruído ou resíduo no tempo t.

Figura 7 - Decomposição: Temperatura Média Diária (°C) - Itapeva/SP 2023-2024



Fonte: Elaborado pelo autor (2024).

Segundo Morettin e Bussab (2017), uma suposição predominante na análise de séries temporais é considerar que a série é estacionária, o que implica que ela evolui ao longo do tempo de forma estocástica em torno de uma média constante. Isso significa que as principais características da série, como a média (μ) e a variância (σ^2), permanecem constantes ao longo do tempo, e a covariância entre os valores só depende da defasagem entre eles, e não do tempo em que foram observados.

O mesmo autor ressalta, que na maioria dos casos, as séries temporais são não estacionárias, conforme Figura 6, onde os requisitos de estacionalidade não são atendidos. Para tratar esses casos, é comum aplicar transformações, como a diferenciação, que é usada para isolar a componente aleatória da série. A diferenciação envolve calcular as variações entre valores sucessivos da série, a partir da equação (16) para modelo aditivo e (17) para multiplicativo (conhecida como *log return*). O objetivo é a retirada da tendência e a sazonalidade, aproximando a série de um comportamento mais estável, próximo de uma distribuição normal, permitindo que seja considerada estacionária.

$$\Delta X_t = X_t - X_{t-1} \tag{16}$$

$$\Delta \log X_t = \log X_t - \log X_{t-1} \tag{17}$$

Onde ΔX_t é chamada de integrada de ordem um, podendo ser chamada de $X_t \sim I(1)$. Na

maioria das situações, é necessário aplicar a diferenciação mais de uma vez, para que se consiga remover todas as componentes de tendência e sazonalidade. Nesse caso, a diferenciação é generalizada para uma ordem n, representando-se como $X_t \sim I(n)$ onde n indica o número de vezes que a diferenciação foi aplicada.

2.3.1.2.2 Tendência

Uma série temporal pode ser modelada com base em duas estruturas principais que influenciam as previsões: a estrutura sistemática e a estrutura aleatória. A estrutura sistemática refere-se aos padrões e tendências consistentes ao longo do tempo, que podem ser identificados e estimados por meio de procedimentos inferenciais, utilizando a amostragem de dados. Por outro lado, a estrutura aleatória abrange eventos imprevisíveis, que ocorrem de maneira aleatória e são estimados por métodos probabilísticos. Embora o foco principal geralmente esteja nos elementos sistemáticos, é importante reconhecer que a aleatoriedade está sempre presente. Por isso, ao construir um modelo de série temporal, é essencial realizar ajustes para minimizar os efeitos dos componentes randômicos, garantindo previsões mais assertivas (Sousa *et al.*, 2021).

Ainda conforme o autor citado, a tendência em uma série temporal é identificada quando há a observação de um comportamento consistente de crescimento ou decrescimento ao longo do tempo, podendo ser descrito pela equação:

$$X_t = T_t + R_t, t = 1, 2, \dots, n \tag{18}$$

Onde X_t representa um ponto específico de n observações em uma série temporal X no tempo (t), T_t a tendência associada e R_t é a variável aleatória.

Morettin e Toloi (2018) destaca que há diversas metodologias, tanto paramétricas quanto não paramétricas, para estimar o componente de tendência em séries temporais. As técnicas paramétricas assumem uma forma funcional específica, como modelos lineares, polinomiais ou exponenciais, e estimam os coeficientes correspondentes. Em contraste, as abordagens não paramétricas, não impõem uma forma funcional definida para a tendência.

Uma abordagem não paramétrica muito utilizado para estimação de tendência é a suavização por média móvel, que consiste em estimar o componente de tendência em um ponto arbitrário t (tempo) a partir da média ponderada dos eventos anteriores e posteriores ao instante

t, representado na equação (19) (Morettin; Toloi, 2018).

$$\widehat{\mathrm{T}}_{\mathrm{t}} = \frac{1}{2m+1} \sum_{j=-m}^{m} X_{\mathrm{t+j}} \mid t = m+1, \dots, n-m \tag{19}$$

Com $\widehat{\mathrm{T}}_{\mathrm{t}}$ representando a estimativa da tendência no instante t, $\mathrm{X}_{\mathrm{t+j}}$ refere-se aos valores da série temporal variando conforme t e m (número de observações desejadas).

2.3.1.2.3 Sazonalidade

A sazonalidade é caracterizada pelo comportamento recorrente de repetitivos padrões em intervalos específicos de tempo, como dias, meses ou anos. Esse comportamento cíclico reflete de eventos periódicos, como estações do ano, feriados ou eventos específicos, podendo ser definida pela equação (20) (Morettin; Toloi, 2018).

$$X_{\mathrm{t}} = S_{\mathrm{t}} + R_{\mathrm{t}}, \mathrm{t} = 1, 2, \dots, \mathrm{n} \tag{20}$$

Onde X_{t} representa um ponto específico de n observações em uma série temporal Y no tempo (t), $\mathrm{S_t}$ a sazonalidade associada e ε_{t} é a variável aleatória.

Morettin e Toloi (2018) cita alguns métodos utilizados para determinação da sazonalidade em uma série temporal, dentre eles, vale destacar o método de médias móveis. Dada uma série temporal X_t, a média móvel simples de ordem k, pode ser calculada pela equação (21).

$$\mathrm{S_t} = X_{\mathrm{t}} - \left[\frac{1}{k} \sum_{i=-\frac{k-1}{2}}^{\frac{k-1}{2}} X_{t+i} \right] \tag{21}$$

Onde $\mathrm{S_t}$ representa a sazonalidade calculada em instante t arbitrário, k é o número de períodos no ciclo sazonal (por exemplo, 12 para dados mensais), e X_{t+i} são os valores da série temporal considerando o ciclo de sazonalidade definido previamente.

2.3.1.2.4 Resíduo (componente aleatório)

Sousa *et al.* (2021) define resíduos em uma análise de séries temporais são as diferenças entre os valores observados da série e os valores ajustados ou previstos por um modelo. Em outras palavras, os resíduos representam a parte não explicada do modelo, que pode ser considerada como erro ou variação aleatória. A fórmula para calcular os resíduos é:

$$R_t = X_t - (T_t + S_t) \tag{22}$$

Analisar os resíduos é fundamental para avaliar a adequação do modelo, pois eles devem ser aleatórios e não apresentar padrões significativos. Isso indica que o modelo está capturando a estrutura da série de forma adequada, conhecido como ruído branco, que refere a uma sequência de variáveis aleatórias que são independentes e identicamente distribuídas (i.i.d.), com média zero e variância constante. Em uma série temporal, o ruído branco é considerado um componente aleatório que não possui autocorrelação, ou seja, as observações em momentos diferentes não têm relação entre si (Sousa *et al.*, 2021).

2.3.1.2.5 Média móveis simples (MMS)

Segundo Morettin e Toloi (2018) a média móvel consiste em estimar o valor de um determinado instante no tempo com base em valores passados, sendo uma técnica fundamental para os métodos subsequentes. Quando se considera uma série estacionária, o cálculo da média móvel pode ser representado pela seguinte equação:

$$\hat{X}_t(h) = \hat{X}_{t-1}(h) + \frac{X_t - X_{t-r}}{r} \tag{23}$$

Acima, $\hat{Z}_t(h)$ é a representação de previsão de h passos da unidade temporal da série ($\forall h > 0$), sendo $X_t = \mu + R_t$. Com r sendo os valores anteriores a serem considerados na previsão.

2.3.1.2.6 Suavização Exponencial

Os métodos de suavização são amplamente utilizados em diversas áreas, incluindo em

call centers, devido à sua simplicidade e eficácia na redução da componente aleatória das séries temporais. A premissa fundamental desses métodos é suavizar o "ruído" nas séries através da utilização de observações passadas, aplicando-lhes pesos que decrescem exponencialmente. Assim, as observações mais recentes recebem maior peso em comparação com as mais antigas. Esse efeito de redução de "ruído" permite que os padrões subjacentes nas séries temporais fiquem mais evidentes.

Um dos métodos mais simples é a Suavização Exponencial Simples (SES), indicada para séries temporais que são localmente constantes, ou seja, que não apresentam tendência ou sazonalidade. Por outro lado, existem métodos mais complexos que consideram essas variações, como a Suavização Exponencial de Holt (SEH) e a Suavização Exponencial de Holt-Winters (SHW), adequados para séries com tendência e sazonalidade (Hyndman; Athanasopoulos, 2021).

Segundo Hyndman e Athanasopoulos (2021), o método SES aplica um suavizador para calcular médias ponderadas das observações, sem considerar componentes de tendência ou sazonalidade. Este método é formalmente representado pela equação (24), onde as observações mais recentes recebem maior peso, enquanto as observações mais antigas perdem relevância conforme se distanciam no tempo.

$$\hat{X}_t(h) = X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2X_{t-2} + \cdots + \alpha(1-\alpha)^nX_{t-n} \tag{24}$$

O componente t representa a unidade de tempo da previsão da componente sistemática X_t conforme já definido em 2.3.1.2.5, n é a quantidade de observações e a é o parâmetro de suavização que varia entre 0 e 1. O parâmetro a depende da influência desejada das observações mais antigas. Um valor maior de a dá mais peso às observações recentes e reduz menos o impacto das observações passadas, enquanto um valor menor de a diminui mais rapidamente o peso das observações antigas, enfatizando mais as observações recentes. A variável h representa a posição passos à frente, com valores maiores que 0.

Segundo Morettin e Toloi (2018), o método SEH consegue estimar tanto a componente sistemática $\hat{X}_t$ e tendência $\hat{T}_t$ da série temporal. As componentes são representadas pelas equações (25), (26) e (27).

$$\bar{X}_t = \alpha X_t + \alpha(1-\alpha)(\bar{X}_{t-1} + \hat{T}_{t-1}) \tag{25}$$

$$\hat{T}_t = \beta(\bar{X}_t - \bar{X}_{t-1}) + (1 - \beta)\hat{T}_{t-1} \quad (26)$$

$$\hat{X}_t(h) = h \times \hat{T}_t \quad (27)$$

A variável $\hat{T}_t$ representa a estimativa da tendência na série temporal, enquanto os coeficientes α e β são os parâmetros de suavização, ambos variando entre 0 e 1. Conforme mencionado anteriormente, o coeficiente α controla o grau de suavização aplicado às observações da série temporal, influenciando no impacto das observações passadas e recentes. O coeficiente β, por sua vez, exerce um efeito semelhante, mas é aplicado especificamente à componente de tendência da série.

Observa-se que a sazonalidade em uma série temporal se refere ao comportamento cíclico de eventos que se repetem em intervalos regulares. Diante disso, torna-se necessário prever séries temporais que considerem tanto a tendência quanto a sazonalidade. Para atender a essa necessidade, utiliza-se o método SHW, que produz estimativas baseadas em três componentes sistemáticas: nível, sazonalidade e tendência. O SHW pode ser aplicado em dois modelos distintos: o aditivo, quando a variação sazonal é constante, e o multiplicativo, quando a variação sazonal é proporcional ao nível da série. O modelo aditivo é representado a partir das equações (28), (29) e (30) (Morettin; Toloi, 2018).

$$\bar{X}_t = \alpha\left(X_t - \hat{S}_{t-p}\right) + (1 - \alpha)\left(\bar{X}_{t-1} + \hat{T}_{t-1}\right) \quad (28)$$

$$\hat{S}_t = \gamma(X_t - \bar{X}_t) + (1 - \gamma)\hat{S}_{t-p} \quad (29)$$

$$\hat{X}_t(h) = \bar{X}_t + h\hat{T}_t + \hat{S}_{t+h-p}\,, h = 1, \ldots, p \quad (30)$$

Diferente dos anteriores, foi adicionado o componente de sazonalidade representado pelo estimador $\hat{S}_{t-p}$, com p equivalente ao período de sazonalidade observado na série, isto é, por exemplo $p = 6$ para apurar picos sazonais de vendas semestralmente. O coeficiente γ, por sua vez, exerce um efeito semelhante aos coeficientes apresentados (α e β), mas é aplicado especificamente à componente de sazonalidade da série.

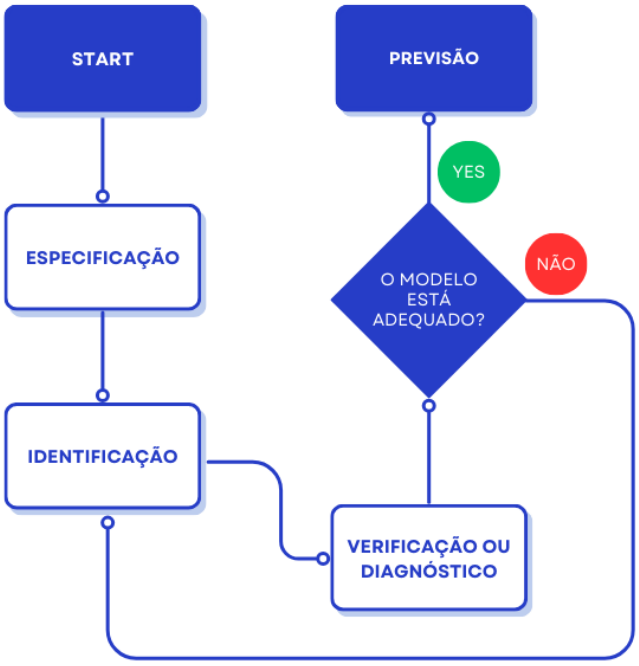
Ainda segundo o mesmo autor, há o modelo multiplicativo, que se pode obter de maneira análoga aditivo, seguido conforme a equação (31), e para previsões $\hat{X}_t(h)$ segue a equação (29).

$$\hat{X}_t(h) = \bar{X}_t + h\hat{T}_t(\hat{S}_{t+h-p}) \, , h = 1, 2, \dots, p \tag{31}$$

2.3.1.2.7 ARIMA

Morettin e Toloi (2018) define ARIMA como um modelo iterativo, em que a melhor modelagem é escolhida a partir dos próprios dados. São compostos por 5 etapas: especificação, identificação, estimação, verificação ou diagnóstico e aplicação (por exemplo previsão), ilustrado na Figura 8.

Figura 8 - Processo Iterativo ARIMA



Fonte: Elaborado pelo autor (2024).

2.3.1.2.8 Especificação

Segundo Morettin e Toloi (2018) a modelagem ARIMA tem sua origem a partir da metodologia de Box-Jenkins de 1976, sendo acrônimo para Autoregressivas Integradas por Médias Móveis, é indicada por meio da notação ARIMA (p, d, q). Dentro dessa estrutura os parâmetros p, d e q, que são todos números inteiros não negativos, representam a quantidade de componentes autorregressivos, a contagem de diferenciações executadas para tornar a série estacionária e o número total de componentes da média móvel, respectivamente. Além disso,

existem variantes particulares da estrutura ARIMA, como o modelo ARMA (p, q), que significa médias móveis autorregressivas, juntamente com o modelo Autorregressivo AR (p) e o modelo de Média Móvel MA (q), ambos relevantes para séries temporais estacionárias caracterizadas por uma contagem nula de diferenças.

O Modelo Autoregressivo de ordem p - AR(p): é representado pela equação (32), onde a variável dependente X_t é definida como uma função linear de seus p valores passados: $X_{t-1}, X_{t-2}, \dots, X_{t-p}$. O parâmetro p indica a quantidade de observações anteriores consideradas no modelo. O termo μ refere-se à média da série, que é igual a zero quando X_t são estacionárias. Os coeficientes $\varphi_1, \varphi_2, \dots, \varphi_p$ são as constantes autoregressivas, geralmente estimadas por métodos como os mínimos quadrados ou máxima verossimilhança. Além disso, a média condicional pode ser expressa como $m = \mu(\varphi_1 - \varphi_2 - \cdots - \varphi_p)$ (Shumway; Stoffer, 2017).

$$X_t = m + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + R_t \tag{32}$$

O Modelo de Médias Móveis de ordem q - MA(q): é representado pela equação (33), onde a variável dependente X_t é definida como a soma do ruído branco atual com os passados. O parâmetro q indica a quantidade de observações anteriores consideradas no modelo. Os coeficientes $\theta_1, \theta_2, \dots, \theta_p$ são os parâmetros de média móvel, geralmente estimadas por métodos como os mínimos quadrados ou máxima verossimilhança (Shumway; Stoffer, 2017).

$$X_t = R_t + \theta_1 R_{t-1} + \theta_2 R_{t-2} + \cdots + \theta_q R_{t-q} \tag{33}$$

O modelo ARMA (p, q) combina os componentes autoregressivos (AR) e de médias móveis (MA), resultando em uma abordagem mais robusta para modelagem de séries temporais. Esse modelo é representado pela união das duas metodologias, conforme indicado na equação (34). O modelo ARIMA (p, d, q), por sua vez, estende o ARMA ao incluir o parâmetro d, que representa o grau de diferenciação necessário para tornar uma série temporal estacionária. A diferenciação remove tendências ou padrões sazonais, conforme discutido no item 2.3.1.2.1, tornando a série adequada para análise (Morettin; Toloi, 2018).

$$X_t = \underbrace{m + \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + R_t}_{} + \underbrace{\theta_1 R_{t-1} + \cdots + \theta_q R_{t-q}}_{} \tag{34}$$

Segundo Morettin e Toloi (2018) o modelo ARIMA pode ser aplicado em séries temporais que, inicialmente, não são estacionárias, mas que se tornam estacionárias após a diferenciação apropriada. Isso amplia significativamente sua aplicabilidade em dados do mundo real, onde as séries frequentemente exibem tendências ou sazonalidades que precisam ser ajustadas para melhorar a precisão das previsões. Dessa forma, o ARIMA é capaz de lidar com padrões mais complexos e dinâmicos em séries temporais, tornando-se uma ferramenta essencial para previsões baseadas em dados históricos.

Apesar da ampla aplicabilidade do modelo ARIMA, ele é classificado como um modelo de memória curta, indicando que ele captura dependências temporais somente entre observações temporalmente próximas. Em um sentido mais específico, o ARIMA é proficiente em modelar séries temporais nas quais a influência de eventos antecedentes diminui rapidamente à medida que o tempo avança. Essa característica surge da presunção do modelo ARIMA de que as correlações entre observações que estão mais distantes temporalmente provavelmente são mínimas, concentrando-se principalmente nas interações entre valores adjacentes ou próximos (Brockwell; Davis, 2002).

De acordo com o mesmo autor citado, o fenômeno de “memória curta” restringe a eficácia do modelo ARIMA em cenários em que eventos ou padrões históricos mantêm um impacto duradouro nas manifestações comportamentais futuras. Para séries temporais que exibem características de memória longa, ou seja, onde eventos antecedentes afetam a série em intervalos temporais mais extensos, torna-se imperativo explorar abordagens alternativas de modelagem, que facilita a modelagem de dependências de longo prazo ou, potencialmente, a implementação de modelos que encapsulam influências estruturais mais profundas, como mudanças de regime ou padrões sazonais intrincados.

2.3.1.2.9 Identificação

A etapa de identificação consiste em identificar as ordens p e q em uma modelagem ARIMA. Porém, antes deve-se verificar a estacionalidade da série temporal. Como já mencionado, na maioria das vezes lidamos com séries não estacionárias, sendo necessário aplicar transformações, como a conversão dos valores em logaritmo ou a diferenciação (d). Após essas transformações, é recomendado utilizar os gráficos da Função de Autocorrelação (FAC) e da Função de Autocorrelação Parcial (FACP) dos valores transformados para

identificação dos parâmetros (Morettin; Bussab, 2017).

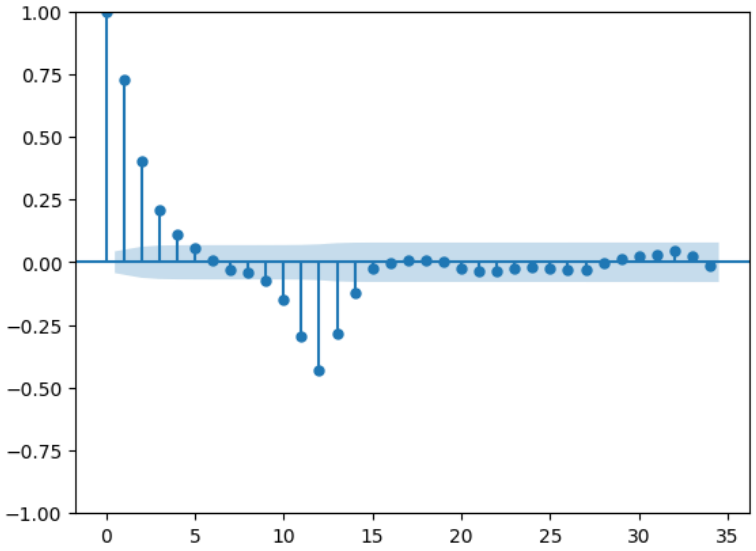
O FAC pode ser definida pela equação (35) e ilustrado pela Figura 9, consiste em medir o grau de correlação serial entre os valores atuais de uma série temporal e seus valores passados em diferentes defasagens (h), chamadas de *lags* a partir da divisão entre a $Cov(X_{t+h}, X_t)$ pela raiz quadrada do produto das variâncias do valor dado um h e $Cor(X_{t+h}, X_t)$ o Coeficiente de Correlação, que é análogo aos cálculos demostrados no item 2.3.1.1 sobre regressões, podendo variar entre -1 a 1, com os extremos indicando um nível maior de correlação negativamente ou positivamente. É uma ferramenta importante para identificar a estrutura de dependência temporal da série, ou seja, o quanto os valores em um determinado ponto no tempo estão correlacionados com os valores anteriores (Morettin; Toloi, 2018).

$$\text{FAC}(h) = \rho(h) = Cor(X_{t+h}, X_t) = \frac{\gamma_h}{\gamma_0} = \frac{Cov(X_{t+h}, X_t)}{Var(X_t)} \tag{35}$$

Onde $-1 \leq \rho(h) \leq 1$ para $h > 1$ e $\rho(h) = 1$ para $h = 0$. Ao considerar uma única série temporal observada, na prática geralmente é utilizado a função de autocorrelação amostral, conforme equação (36).

$$\hat{\rho}(h) = \frac{\hat{\gamma}_h}{\hat{\gamma}_0} = \frac{\frac{1}{n}\sum_{t=1}^{n-h}[(x_t - \bar{x})(x_{t+h} - \bar{x})]}{\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2} \tag{36}$$

Figura 9 - Exemplificação Gráfico de Autocorrelação

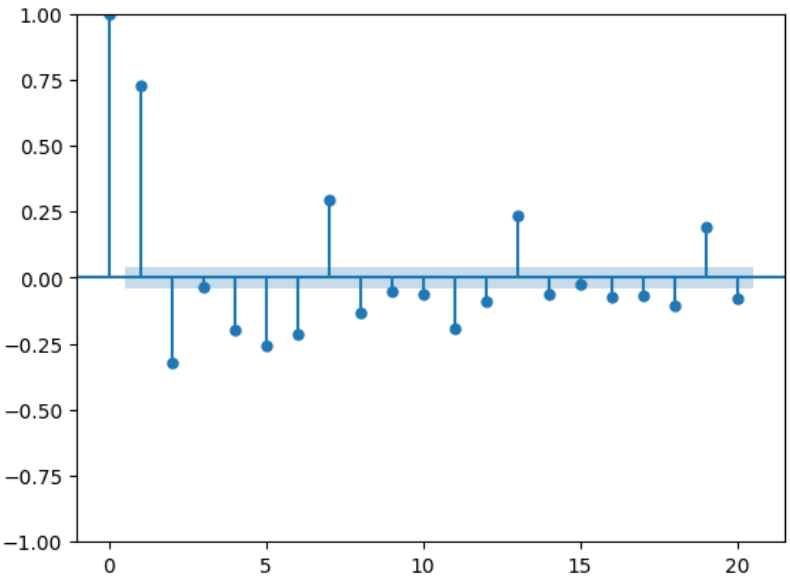


Fonte: Elaborado pelo autor (2024).

Segundo Hyndman e Athanasopoulos (2021), a Função Autocorrelação Parcial (FACP) mede a correlação direta de valores de uma série temporal X_t e seus valores passados X_{t-k} com $k = 1, 2, \dots, n$, eliminando os efeitos intermediários, isto é, a influência de valores situados entre os dois pontos que estão sendo comparados. Em outras palavras, ao considerar três eventos, A, B e C, a FACP avalia a correlação entre A e C, desconsiderando o impacto de B, mesmo que haja uma correlação entre A e B ou entre B e C. A função de correlação parcial é descrita pela equação (37) e ilustrada na Figura 10.

$$FACP(h) = \pi(h) = Corr(X_t, X_{t-k} | X_{t-1}, X_{t-2}, \dots, X_{t-k+1}) \quad (37)$$

Figura 10 - Exemplificação Gráfico de Autocorrelação Parcial



Fonte: Elaborado pelo autor (2024).

Esses gráficos ponderações para escolha das ordens p e q dos modelos autoregressivos (AR) e de média móvel (MA), ajudando a visualizar padrões temporais, tendências ou sazonalidades na série.

Ao trabalhar com os gráficos de FAC e FACP, é comum interpretar os resultados utilizando intervalos de confiança baseados na distribuição normal. Isso se deve ao fato de que, ao aproximar a série temporal de uma condição estacionária, espera-se que os dados exibam um comportamento aproximadamente gaussiano. Normalmente, utiliza-se um intervalo de confiança de 95%, o que permite definir os limites a partir da equação (38), onde N representa

o número de observações da série. Na interpretação dos gráficos, as correlações que se encontram dentro desse intervalo de confiança são consideradas ruídos brancos, ou seja, tendem a zero, indicando a ausência de correlação significativa (Hyndman; Athanasopoulos, 2021).

$$\pm \frac{1{,}96}{\sqrt{N}} \tag{38}$$

Segundo Morettin e Toloi (2018) pode-se obter algumas interpretações das funções de autocorrelação conforme descrito abaixo:

- Para dados que seguem um modelo AR (p), a FAC apresenta um comportamento infinito com decaimento exponencial, enquanto a FACP exibe um ou mais cortes fora do intervalo de confiança.
- Para dados que seguem um modelo MA (q), o comportamento é inverso ao do AR(p), ou seja, a FAC exibe um ou mais cortes fora do intervalo de confiança, enquanto a FACP decai exponencialmente.
- Para dados que seguem um modelo ARMA (p, q), observa-se uma combinação entre o decaimento exponencial e os cortes em ambos os gráficos. Os cortes na FAC sugerem os candidatos para a ordem q, e os cortes na FACP indicam a ordem p.

Além da análise gráfica, os autores sugerem avaliar os índices Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC), conforme equações (39), (40) e (41).

$$AIC = -2\log(L) + 2(p + q + k + 1) \tag{39}$$

$$\mathrm{AIC_c} = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2} \tag{40}$$

$$BIC = AIC + [\log(N) - 2][p + q + k + 1] \tag{41}$$

Aqui, L representa a verossimilhança do modelo, k o número de parâmetros, enquanto p e q são as ordens definidas graficamente a partir dos gráficos de autocorrelação. Já N corresponde ao número de observações no conjunto de dados. Modelos com menor valor de AIC são preferíveis, pois indicam um melhor equilíbrio entre o ajuste do modelo e sua complexidade. Por outro lado, modelos com menor valor de BIC são considerados melhores, pois sugerem uma maior probabilidade dos dados, ajustando-se ao modelo com uma

penalização adicional pela complexidade, favorecendo modelos mais simples.

2.3.1.2.10 Estimação

Segundo Vasconcellos e Alves (2000) a estimação dos parâmetros $\emptyset$ *e* θ comumente é utilizado alguns métodos de ajustes para sistemas lineares e/ou não-lineares, como:

- Métodos dos Momentos;
- Estimadores Mínimos Quadrados;
- Estimadores de máxima verossimilhança

Em sistemas não-lineares, especialmente em casos que envolvem modelagem com o componente MA(q) (Média Móvel), a estimação dos parâmetros torna-se extremamente complexa e computacionalmente custosa. Devido à natureza não-linear desses modelos e à dificuldade de obter soluções analíticas, a utilização de softwares especializados torna-se essencial para realizar a estimação de forma eficiente e precisa. Esses programas são projetados para lidar com a complexidade e os requisitos computacionais elevados, permitindo o ajuste adequado dos modelos MA(q).

2.3.1.2.11 Verificação ou Diagnóstico

Morettin e Toloi (2018) e Hyndman e Athanasopoulos (2021) apontam alguns que podem ser aplicados com objetivo de verificar os ajustes dos modelos propostos aos dados. Esses testes avaliam a adequação do modelo, sendo possível identificar oportunidades de melhoria na modelagem ou a necessidade de considerar modelos alternativos. Entre os testes recomendados, destacam-se o Teste de Box-Pierce-Ljung, que verifica se os resíduos do modelo são independentes, o Teste de Autocorrelação Cruzado, que verifica a existência de novos termos de médias móveis a serem adicionados no modelo, o Teste de Dickey-Fuller Aumentado (ADF), que avalia a estacionalidade da série temporal, e os testes de normalidade, como o de Shapiro-Wilk e Jarque-Bera, que verificam se os resíduos seguem uma distribuição normal. A aplicação desses testes permite uma análise mais robusta e a identificação de possíveis ajustes para melhorar a qualidade do modelo.

No escopo do trabalho não será abordado com detalhes o funcionamento de cada um dos testes, visto que nos diversos artigos estudados sobre aplicação de ARIMA em *Call Center*, os autores utilizaram em grande maioria as análises presentes no item 2.3.1.2.9 para definir

potenciais ordens da modelagem ARIMA, gráficos FAC dos resíduos após previsão do modelo e gráficos (*Q-Q plot* e histograma) para verificar a normalidade dos dados. Os resultados e conclusões sobre esses artigos vai ser abordado nos itens de revisão de literatura.

Segundo Morettin e Toloi (2018) as conclusões sobre verificação da normalidade via gráfico são:

- Gráfico FAC e FACP dos resíduos: Os dados devem estar dentro do intervalo de confiança de 95%;
- Q-Q *plot*: O funcionamento é a partir da comparação dos quantis dos dados da amostra observada no eixo y e os quantis teóricos respeitando uma distribuição normal no eixo x, em que o ajustamento linear entre eles indica a aproximação dos dados a uma distribuição gaussiana;
- Histograma: Os dados aproximam-se de uma distribuição normal quando, visualmente, a distribuição de frequência dos dados assume um formato de sino simétrico, conhecido como curva normal ou curva de Gauss. Essa forma indica que a maior parte dos valores está concentrada ao redor da média, com uma diminuição gradual da frequência à medida que se afastam da média em ambas as direções.

2.3.1.2.12 Previsão

Na maioria das vezes utiliza-se os estudos com séries temporais para previsão de algum evento, para previsões por exemplo do objeto de estudo do trabalho, a previsão da volumetria de chamadas em um *call center*, não é diferente. Considerando a modelagem ARIMA, uma das maneiras de representação matemática é denominada equação das diferenças, representada na equação (42), em que $h > 1$, sendo a defasagem futura desejada para previsão. O principal balizador de qualidade da previsão é a minimização das diferenças (erro) entre valores reais e preditos, com isso utiliza-se dos mesmos indicadores já citados no item 2.3.1.1.1, em que foi abordado RMSE e MAPE para modelagens envolvendo regressões, que podem ser utilizados de forma análoga aqui para valores X_{t+h} da série temporal (Morettin; Toloi, 2018).

$$X_{t+h} = \varphi_1 X_{t+h-1} + \cdots + \varphi_{p+d} X_{t+h-p-d} + R_{t+h} - \theta_1 R_{t+h-1} - \cdots - \theta_q R_{t+h-q} \quad (42)$$

2.3.2 Inteligência Artificial

A inteligência artificial (IA) surge como uma força transformadora, permeando cada vez mais a sociedade hiper conectada e moldando o cenário socioeconômico global. Segundo Kaufman (2022), a IA já se tornou a tecnologia de propósito geral do século XXI, impulsionando a redução de custos com dados e alavancando a economia. Apesar da crescente presença da IA em áreas como finanças, saúde e entretenimento, é crucial lembrar que, por trás dos algoritmos, reside a subjetividade humana. A construção dos modelos, a definição dos parâmetros e a seleção do domínio da aplicação carregam os valores e preconceitos humanos, tornando a IA uma ferramenta social e humana, com efeitos diretamente ligados à forma como ela é usada.

Lima, Pinheiro e Santos (2014) define a IA como um conjunto de procedimentos computacionais que simulam a inteligência humana, buscando replicar comportamentos inteligentes como a resolução de problemas, a aquisição de conhecimento e o reconhecimento de padrões. O desenvolvimento da IA se deu a partir de diversas áreas de estudo, incluindo a matemática, a lógica e a neurociência. As diferentes abordagens para IA, como a simbólica, a conexionista e a evolucionária, refletem a complexidade do conceito de inteligência e a busca por diferentes caminhos para replicá-la em máquinas.

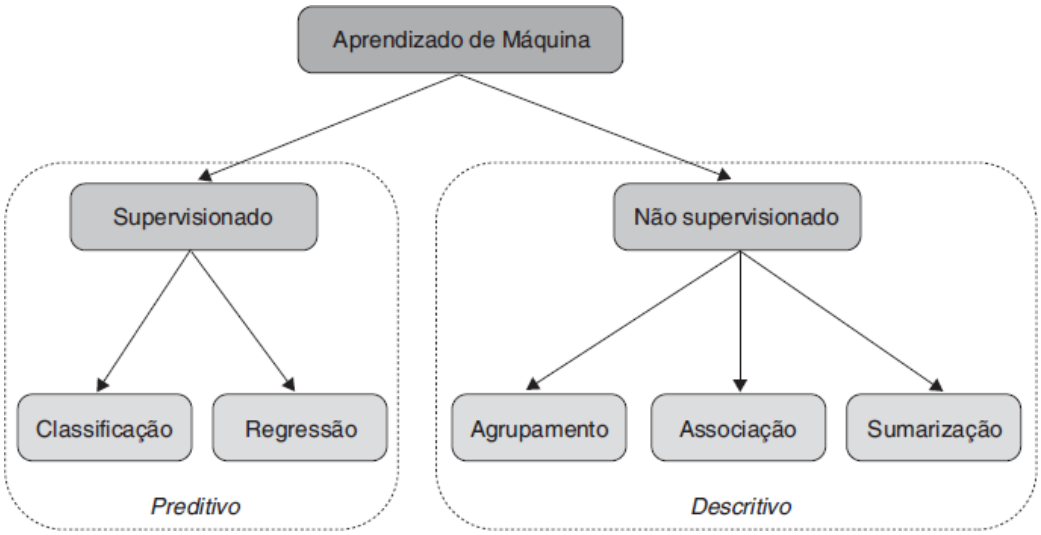
Russell e Norvig (2022), por sua vez, aborda a IA como um campo universal que busca construir entidades inteligentes, capazes de agir de forma eficaz em diversas situações. A definição de inteligência, nesse contexto, se divide em duas vertentes: a similaridade com o desempenho humano e a racionalidade. As diferentes abordagens da IA, como o teste de Turing, a modelagem cognitiva e a ação racional, representam diferentes perspectivas sobre o que define a inteligência e como ela pode ser implementada em sistemas computacionais. A preocupação com o alinhamento de valores, ou seja, garantir que a IA atenda aos objetivos humanos, emerge como um desafio crucial para o futuro da IA, uma vez que a crescente capacidade da IA exige mecanismos para garantir que sua inteligência seja direcionada para o bem da humanidade.

Faceli *et al.* (2021) traça um panorama histórico da IA mostrando como ela evoluiu de programas rigidamente programados para sistemas mais sofisticados e autônomos, baseados em Aprendizado de Máquina (AM). Essa evolução se deu impulsionada pelo aumento exponencial de dados e pela necessidade de ferramentas mais eficazes para processá-los. O AM, por sua vez, permitiu a resolução de problemas complexos, antes intransponíveis, e a criação de aplicações inovadoras em áreas como saúde, segurança e transporte, como os carros autônomos.

2.3.3 Aprendizagem de Máquina para Séries Temporais

A AM é considerada uma subárea crucial da IA. O AM se destaca por sua capacidade de aplicar técnicas para descrever ou realizar a predição de eventos. A análise preditiva, um dos ramos do AM, utiliza modelos que extraem características de uma base de dados, permitindo a previsão de eventos futuros. Esse processo se denomina aprendizagem supervisionada, pois depende de dados rotulados para aprender padrões e realizar previsões. Por outro lado, a análise descritiva foca em entender os padrões existentes em determinado grupo de dados, sem necessariamente prever eventos futuros. Essa análise busca identificar tendências, agrupamentos e outras características intrínsecas aos dados, oferecendo insights importantes para tomada de decisão e compreensão dos comportamentos (Faceli *et al.*, 2021).

Figura 11 - Divisão de aprendizado preditivo e descritivo



Fonte: Faceli *et al.* (2021).

Como a aplicação do trabalho para previsão de volume de chamadas em *call center* é uma abordagem supervisionado, utiliza-se a seguir um modelo para exemplificação e entendimento.

2.3.3.1 Modelos Supervisionados

Um modelo supervisionado é definido por um conjunto de observações que contém

variáveis preditoras $x_1, x_2, \dots, x_n$ e uma variável alvo y, que pode ser quantitativa, no caso de previsões, ou qualitativa, para modelagem de classificação. No contexto da previsão de volumetria de chamadas em *call centers*, o interesse está em uma previsão quantitativa, onde a variável y representa a quantidade de chamadas, enquanto o conjunto de características dos registros são as preditoras (Sicsú; Samartini; Barth, 2023).

Tabela 1 - Estrutura modelo AM supervisionado

Preditoras				Alvo
x_1	x_2	...	x_m	y
x_{11}	x_{12}	...	x_{1m}	y_1
x_{12}	x_{22}	...	x_{2m}	y_2
⋮	⋮	⋱	⋮	⋮
x_{1n}	x_{2n}	...	x_{nm}	y_3

Fonte: Adaptado (Sicsú, Samartini e Barth) ((2023))

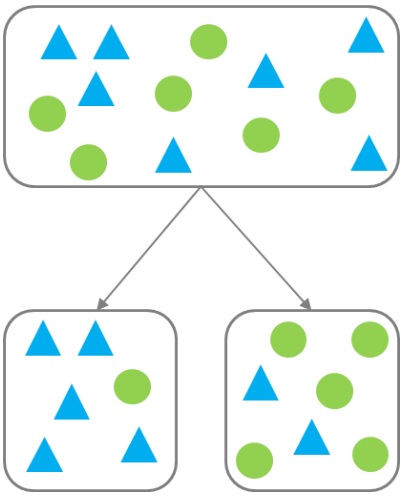
Ainda segundo o autor citado, a variável y está em função $x_1, x_2, \dots, x_n$, conforme equação (43). Dentre as técnicas de modelagem preditiva supervisionada, as principais são: regressões, árvores de regressão, florestas aleatórias, *Gradient Boosting* e *Support Vector Regression*.

$$y = f(x_1, x_2, x_3) \tag{43}$$

2.3.3.2 Arvores de Regressão

Para entendimento do funcionamento de árvores de regressão, primeiro precisa-se entender conceitualmente o funcionamento de uma árvore de decisão. Para Sicsú, Samartini e Barth (2023) uma árvore de decisão, conforme a Figura 12 é uma modelagem simples e de fácil assimilação, que pode ser utilizada para problemas de classificação e regressão.

Figura 12 - Arvore de decisão



Fonte: Elaborado pelo autor (2024).

Uma árvore de decisão é uma estrutura que organiza a derivação de subgrupos com base nos atributos de um conjunto de dados ou em critérios previamente estabelecidos. Cada subgrupo é representado por nós: o nó inicial é chamado de nó raiz, os nós intermediários são os nós filhos, e os nós terminais, que não se ramificam mais, são chamados de nós folha. São esses nós folha que determinam os possíveis resultados da modelagem (Sicsú; Samartini; Barth, 2023).

Segundo Géron (2021), à medida que a árvore de decisão cresce com a criação de mais nós, cada nível de nós é denominado profundidade. Esse parâmetro define a qualidade da segmentação: em problemas de classificação, ele determina a pureza das classes em cada nó, e, em problemas de regressão, reflete a proximidade dos valores de previsão em relação à realidade.

O autor ainda cita que quanto maior a profundidade, mais detalhada e pura será a segmentação, porém isso pode levar ao risco de *overfitting*, que é o fenômeno de sobreajuste aos dados de treinamento. Nesse caso, o modelo se torna excessivamente adaptado às nuances do conjunto de treinamento, podendo capturar até mesmo ruídos ou padrões irrelevantes, o que dificulta a generalização para novos dados reais, isto é, o modelo tem um bom ajuste com dados de treino, mas falha ao tentar fazer previsões em dados desconhecidos.

Para a avaliação dos nós em uma árvore de decisão, podem ser utilizados diferentes indicadores, como o índice Gini, a entropia, erros de previsão ou até testes estatísticos. Em modelos de classificação, o índice Gini e a entropia são comumente usados para medir a impureza dos nós, ajudando a determinar as melhores divisões. Já em modelos de regressão, os

erros de previsão são mais frequentemente utilizados, como o erro quadrático médio (MSE), representado na (44) (Sicsú, Samartini; Barth, 2023).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{44}$$

Os modelos de árvore de decisão são abordagens que não dependem de testes prévios para serem aplicados, nem de definições claras sobre o comportamento esperado dos dados, como linearidade ou algum tipo de distribuição estatística específica. À medida que as iterações ocorrem e a profundidade da árvore aumenta, a modelagem vai se ajustando aos dados de forma adaptativa. No entanto, é crucial definir uma condição de parada ou um limite para as iterações, evitando que o modelo se ajuste demais aos dados de treinamento, o que pode resultar em *overfitting*, como já mencionado. Por isso, Sicsú, Samartini e Barth (2023) apresenta duas abordagens possíveis, sendo elas: pré-poda e pós-poda.

A pré-poda é aplicada antes da execução da modelagem, com base em definições empíricas feitas pelos analistas, como o limite mínimo de observações por agrupamento ou critérios de parada para o particionamento dos nós. O autor cita que geralmente o uso da pós-poda é mais utilizada pelos analistas, ela envolve executar a modelagem completamente e, em seguida, remover nós que não contribuem significativamente para o desempenho do modelo. Apesar, da pós-poda ser frequentemente mais utilizada, em casos de amostras grandes, pode vir a ser mais eficiente validar o modelo com a pré-poda, visto que, demanda menor processamento computacional.

Em modelos de árvores de regressão, Faceli *et al.* (2021) apresentam alguns métodos utilizados para a divisão dos nós. É mais comum o uso de regras que comparam valores com intervalos estabelecidos, priorizando as decisões que minimizam o erro quadrático médio (MSE).

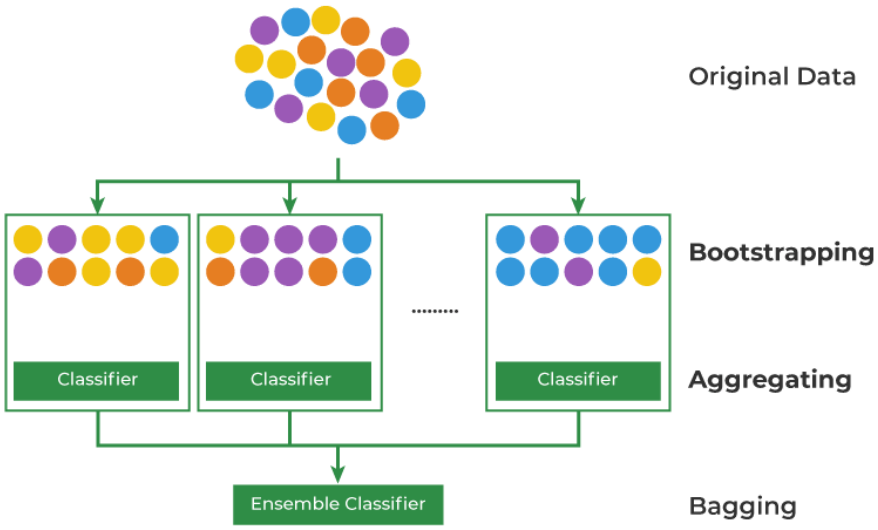
2.3.3.3 Modelagens de Aprendizado Ensemble

O aprendizado ensemble (agrupamento) consiste em modelagens que combinam múltiplos preditores, podendo ser algoritmos diferentes ou de um único submetido a treinamento em diferentes amostras aleatórias da base de treinamento. No final, as predições de cada modelo são consolidadas para formar um resultado. Essa abordagem permite extrair as

melhores características de cada modelagem individual, resultando em um desempenho superior ao de um preditor isolado. No escopo deste trabalho, serão abordados dois métodos de aprendizado ensemble: a floresta aleatória (*random forest*) e o *gradient boosting* (Géron, 2021).

O modelo de floresta aleatória é baseado na técnica de *Bagging* (abreviação de *bootstrap aggregating*) para a segmentação das amostras que serão aplicadas aos diversos preditores de árvores de decisão. Pode-se visualizar na Figura 12 que uma amostra de treinamento, as observações são separadas e agregadas aleatoriamente para realizar uma classificação ou regressão. Neste processo, uma observação pode alocada em várias amostras de diferentes preditores, permitindo que o modelo capture as variações e possa garantir a robustez nas predições (Géron, 2021).

Figura 13 - Classificação *Bagging*



Fonte: Dey (2024)

O processo de previsão em uma floresta aleatória é simples e objetivo. Após a separação e agregação das amostras, várias modelagens de árvore de decisão são aplicadas. O resultado é definido pela agregação das predições dos modelos, utilizando uma abordagem estatística. Em uma modelagem de regressão, por exemplo, a previsão é feita a partir da média dos valores previstos por cada modelo. Esse método de combinação permite que a floresta aleatória obtenha resultados mais robustos e precisos, minimizando a variância e mantendo o viés se comparado com as modelagens individuais (Géron, 2021).

Os modelos de *boosting* são semelhantes à modelagem *bagging*, pois ambos envolvem o agrupamento de vários preditores para melhorar o resultado. No entanto, enquanto o *bagging*

treina os modelos de forma independente e simultânea, o *boosting* executa os algoritmos sequencialmente, de modo que cada novo preditor é ajustado para corrigir os erros dos anteriores, resultando em uma melhoria contínua e performance superiores à das modelagens *bagging*. Os principais são *AdaBoost* (*Adaptive Boosting*), *Gradiente Boosting* e *XGBoost* (*Extreme Gradient Boosting*) (Sicsú; Samartini; Barth, 2023).

Para detalhamento, devido ao escopo do trabalho, o foco será direcionado aos modelos de *Gradient Boosting* e ao *XGBoost*. O *Gradient Boosting* é um modelo que se utiliza de árvores de decisão para classificações ou regressões, a partir da minimização da função de perda definida do modelo, geralmente é para regressões é utilizado a soma dos quadrados dos resíduos, expressa na equação (45) (Sicsú; Samartini; Barth, 2023).

$$\text{SSE } (sum\ os\ square\ errors) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{45}$$

Segundo Sicsú, Samartini e Barth (2023) o modelo *Gradient Boosting c*omumente utiliza árvores de regressão em uma amostra aleatória, com variação de 3 a 6 partições. A cada iteração do treinamento, a função de perda é ajustada para incluir os erros das previsões anteriores, e a variável-alvo é redefinida pela diferença quadrática entre o valor real e a previsão anterior. Em caso em que a função de erros aumenta, há a exigência de maior esforço para melhorar as previsões.

Além disso, é aplicada uma taxa de aprendizado (learning rate), que controla o impacto de cada nova árvore no modelo final, com uma escala que varia de 0 a 1. Valores próximos de zero garantem que as correções sejam feitas de forma mais lenta e gradual. Quanto menor o valor da taxa de aprendizado, melhor pode ser o desempenho do modelo, já que ele se ajusta de maneira mais precisa, porém, isso também aumenta o tempo e o custo de processamento, uma vez que mais iterações serão necessárias para alcançar a previsão final, que é obtida pela última árvore de regressão.

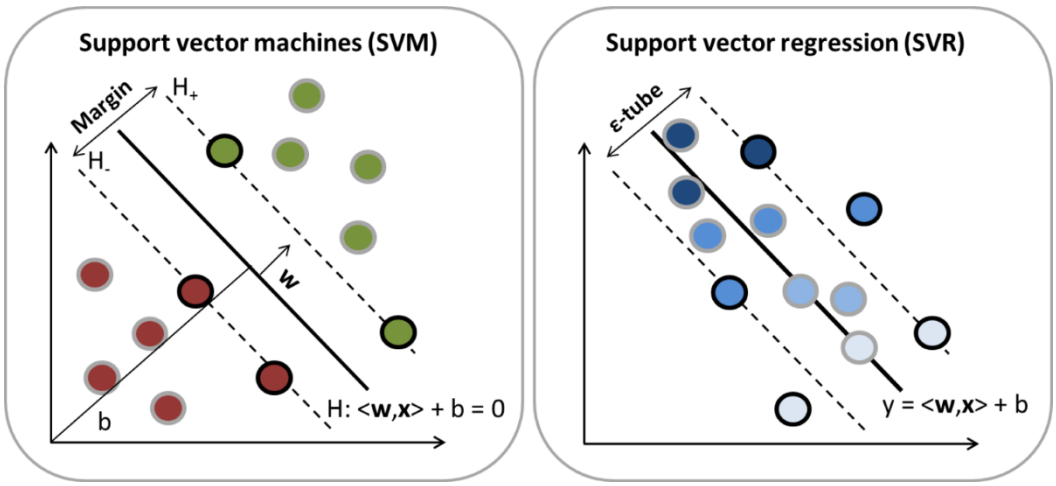
O *XGBoost*, segundo Chen e Guestrin (2016) é um aprimoramento da modelagem *gradiente boosting*, isto é, tem o mesmo funcionamento conceitual, porém com processamento mais rápido, versatilidade de modelos que pode ser aplicado e a possibilidade de aplicar diversos hiper parâmetros com objetivo de melhorar a performance do modelo, tratamento de valores ausentes, definição da profundidade das árvores manualmente, entre outros benefícios.

2.3.3.4 Support Vector Regression (SVR)

Segundo Géron (2021) o *Support Vector Regression* (SVR), ou Regressão de Vetores de Suporte, é derivado do *Support Vector Machine* (SVM), uma técnica amplamente utilizada para modelagem de classificação, o SVR é aplicado em problemas de regressão, embora o conceito básico seja semelhante em ambos.

No SVR, assim como no SVM, existe o uso de um *kernel*, que permite a modelagem de dados em espaços de maior dimensão, facilitando a separação ou ajuste dos dados. No caso da classificação, o modelo tem o objetivo de encontrar um hiperplano que separe as diferentes classes com a maior margem possível. Já no SVR, o objetivo é encontrar uma região (“rua”) com largura de ε (*epsilon*), dentro da qual os dados mais próximos e dentro dessa região são considerados aceitáveis. Já os pontos localizados exatamente sobre as linhas da margem denominados vetor de suporte, sendo os responsáveis por definir posição e largura do hiperplano (Géron, 2021).

Figura 14 - Representação SVM e SVR

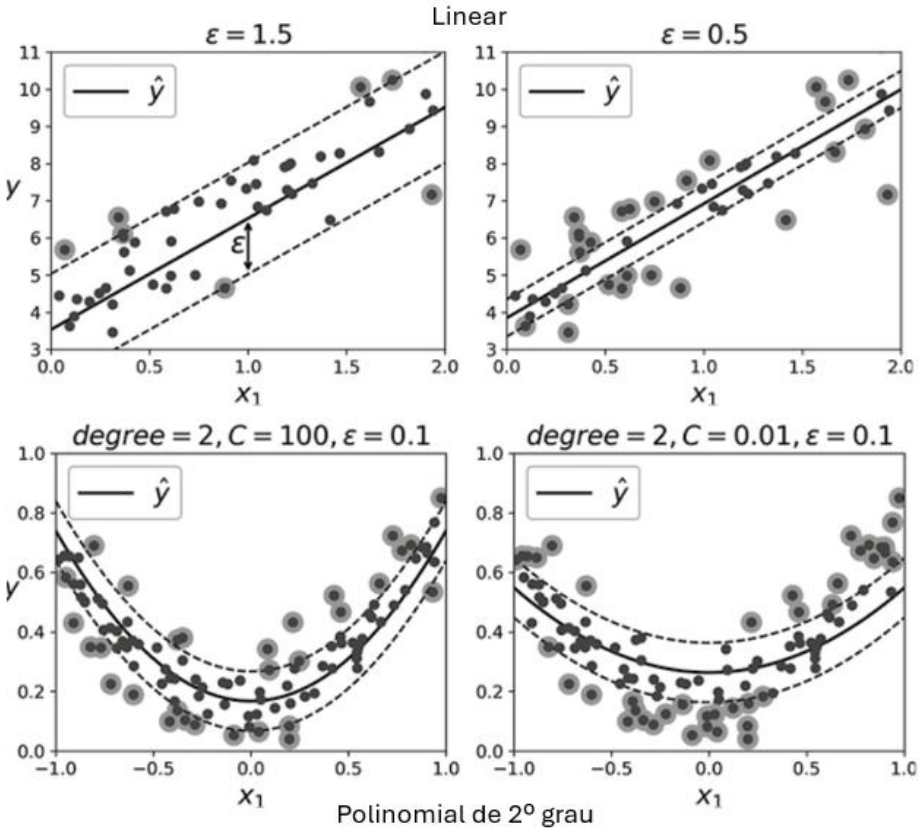


Fonte: Rodríguez-Pérez e Bajorath (2022)

O *Kernel* são os ajustes possíveis, podem ser utilizado o linear, polinomial ou de uma função radial gaussiana, a depender do comportamento do conjunto de dados conforme é ilustrado na . A largura das margens pode ser mais larga ou mais estreita. Margens mais largas têm menor sensibilidade a outliers, mas podem resultar em um desempenho inferior do modelo. Por outro lado, margens mais rígidas (estreitas) tendem a ser mais sensíveis aos outliers, o que pode levar a um ajuste excessivo (*overfitting*) (Géron, 2021).

Por isso, Géron (2021) recomenda o uso de margens suaves, que equilibram esses dois extremos a partir de um parâmetro C. Essa abordagem permite um bom ajuste entre a robustez contra outliers e a capacidade do modelo de generalizar bem os dados. Assim, margens suaves oferecem um compromisso ideal, trazendo melhor desempenho na maioria dos casos.

Figura 15 - Exemplificação *Kernel* Linear e Polinomial



Fonte: Adaptado de Géron (2021).

2.3.3.5 Métricas de Avaliação AM

Para avaliações de modelos de aprendizagem de máquina, de acordo com os autores citados nos tópicos anteriores, Faceli *et al.* (2021), Sicsú, Samartini e Barth (2023) e Géron (2021), as principais métricas para problemas de regressão geralmente são o RMSE, MAPE e MAE, que mede as discrepâncias entre valores reais e previstos, já abordados no item 2.3.1.1.1, sobre métricas de regressão.

2.4 REVISÃO DA LITERATURA

Neste capítulo concentra em citar trabalhos com a aplicação das metodologias apresentadas no segmento de previsão de volumetria de chamadas em *call center*.

Os autores Bouzada e Saliby (2009) abordaram os modelos de regressão, em especial a regressão múltipla por entenderem que é um método de fácil aplicação no contexto de *call centers*. Eles citam que existem vários fatores que influenciam a volumetria de chamadas, como feriados, campanhas de marketing, lançamentos de novos produtos e eventos especiais pré-definidos. Além disso, consideram o dia da semana e datas específicas que variam conforme o produto, como vencimento de faturas e eventos de liquidação como informações cruciais para previsão, com isso o estudo com regressões múltiplas oferece uma análise abrangente, integrando múltiplas variáveis para prever a demanda de forma mais precisa e eficiente.

Bouzada e Saliby (2009) aplicou a regressão múltipla considerando o dia da semana, se é feriado, e a quantidade de contas em diferentes estágios de vencimento (dias antes e após o vencimento) e obteve o valor de R^2 ajustado de 84%, indicando que 84% da variabilidade da demanda de ligações pode ser explicada pelo modelo. Os autores entenderam que foi um bom ajuste do modelo aos dados, demonstrando que as variáveis independentes selecionadas são capazes de explicar grande parte da variabilidade da demanda.

O autor aponta que, apesar de um bom ajuste do modelo indicado pelo coeficiente de determinação (R^2), o MAPE de 10,98% revela que ainda há margem para melhorar a precisão das previsões. Indicando um comportamento crítico do modelo em ambientes onde a demanda é dinâmica e sujeita a variações inesperadas, impossibilitando a previsão satisfatória na presença de eventos atípicos, como feriados prolongados ou datas com comportamentos fora do padrão. Esses tipos de eventos exigem a intervenção de especialistas para ajustar as previsões e garantir que o modelo incorpore essas variações inesperadas de maneira adequada.

Por outro lado, os autores Chanbunkaew e Tharmmaphornphilas (2018) utilizaram métodos tradicionais de séries temporais para previsão de volume de chamadas em uma central de atendimento comercial em um grande banco na Tailândia. O foco do trabalho foi estudar alguns métodos de previsão, como Médias Móveis Simples (MMS), Suavização Exponencial de Holt-Winters (SHW) e ARIMA utilizando a variável de data e volume de chamadas para previsões mensais e diárias, com o objetivo de melhorar a performance dos modelos atuais utilizados no banco. Em ambos, foi utilizado o período de 2011 a 2015 como insumo para previsões do ano de 2016.

Para o modelo MMS, os autores adotaram o parâmetro $r = 3$, utilizando os três valores anteriores para previsões com defasagens h meses ou dias a frente. Para modelagem de SHW

foi utilizado valor de α igual a 1 para garantir maior peso nos valores recentes, minimizando a volatilidade de valores distantes. Para modelagem ARIMA foi configurado com os componentes autoregressivos AR (1) e AR (12) e com ordens d e q iguais a zero para ambos. A escolha se baseou a partir da análise dos gráficos de autocorrelações, onde observaram um comportamento com decaimento exponencial para o gráfico FAC e cortes claros no gráfico FACP em 1 e 12, enquanto o restante dos *lags* e o gráfico FAC dos resíduos permaneceram dentro no intervalo de significância de 95%.

Para previsões mensais, a modelagem ARIMA apresentou melhor ajuste aos dados, com MAPE de 3,47%, seguido pelo método MMS, com 4,31%, e, na última colocação a metodologia SHW, que retornou MAPE de 4,62%. Com as previsões diárias, os autores utilizaram a média mensal para estimativas, resultando em valores de 5,52% a 16,82% de MAPE. Esses resultados superaram os modelos atuais do banco, conforme destaque dos autores. Além disso, eles apontam que as flutuações nas previsões diárias podem ser explicadas por problemas recorrentes nos sistemas do *call center* do banco ocorridas em abril de 2016, que impactou na estimação médias para aquele período (Chanbunkaew; Tharmmaphornphilas, 2018).

Ao comparar isoladamente a modelagem utilizada pelo primeiro autor citado, é possível observar que Chanbunkaew e Tharmmaphornphilas (2018) obteve resultados melhores ao comparar o indicador MAPE entre os modelos. Isso indica um melhor ajuste dos dados com modelos de séries temporais. Entretanto, para comparações mais justas, há a necessidade de aplicação dos modelos em condições semelhantes. Além disso, outros estudos, como de Rausch e Albrecht (2020), adotam a aplicação de aprendizagem de máquina, abordando comparações dos resultados com as modelagens clássicas citadas.

Rausch e Albrecht (2020) propuseram estudar a acurácia de vários modelos, incluindo regressões, modelos clássicos de séries temporais e aprendizado de máquina, para previsão de chamadas em um *call center* de um varejista com atuação online. A proposta comtempla também, a comparação de performance dos modelos considerando a segmentação dos dados em diferentes tempos previstos para entrega dos produtos.

Os autores consideraram o período de 2 de janeiro de 2016 a 7 de maio de 2019 e, após algumas exclusões devido a dados incorretos, conseguiram analisar 31.050 observações no período mencionado. Essas observações contemplam chamadas telefônicas ou pedidos via e-mail solicitando suporte ao varejista. Embora o artigo não detalhe explicitamente os valores das componentes de cada modelo, como os parâmetros α para as modelagens de suavização exponencial ou as ordens da modelagem ARIMA, é interessante focar nas comparações entre

os modelos e nos resultados obtidos.

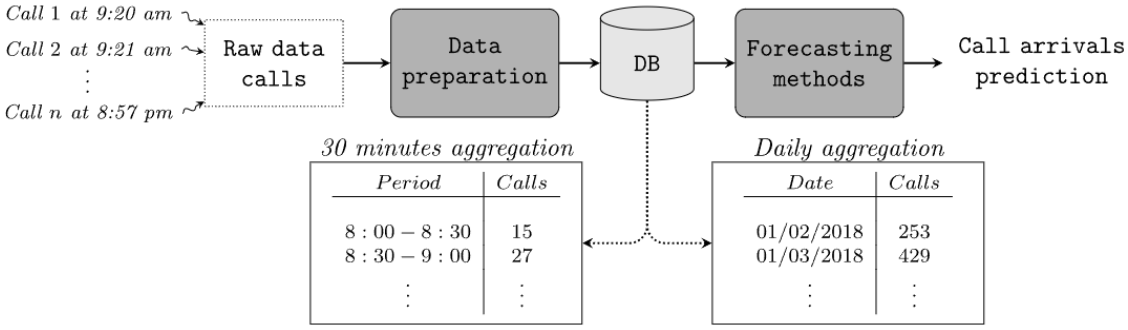
Para a comparação dos resultados, os autores concluíram que a abordagem de aprendizado de máquina, especificamente *Random Forest*, apresentou os melhores resultados em ambos os cenários, mesmo com as segmentações por tempo de entrega. Em seguida, os modelos de regressões múltiplas, que consideraram distribuições de probabilidades, e os modelos ARIMA obtiveram desempenhos inferiores. Por último, as variações utilizando Suavizações Exponenciais apresentaram os piores resultados (Russell; Norvig, 2022).

Além do método de *Random Forest* proposto por Rausch e Albrecht (2020), os autores Chacón *et al.* (2023) propõem outros métodos além dos clássicos, como *Gradient Boosting* e modelos de *Deep Learning*. Foram utilizados dados de janeiro a outubro de 2018 de uma seguradora dos Estados Unidos, com ligações diárias registradas de segunda a domingo, com atendimento das 6:00 às 22:00.

Para as previsões, foi seguido o processo ilustrado na Figura 16 para agregação dos dados. A partir disso, foram aplicados os modelos de SHW aditivo e multiplicativo, SARIMA, *xGBoost* (como método de *Gradient boosting*), além de dois modelos de *Deep Learning*. Para o intervalo intradiário os dois modelos de *Deep Learning* obtiveram os melhores resultados, com *xGBoost* ficando em terceiro lugar. Já para resultados diários, *xGBoost* obteve o melhor desempenho, seguido pelo SHW multiplicativo e SARIMA, respectivamente (Chacón *et al.*, 2023).

Ainda conforme os autores citados, à medida que a granularidade dos dados aumenta temporalmente, há a intensificação da variabilidade sazonal. Isso faz com que os modelos de *Deep Learning* tenham dificuldade em capturar esses padrões, resultando em um desempenho abaixo quando comparados aos modelos de aprendizado de máquina e clássicos. Por outro lado, as modelagens clássicas tendem a ter melhor desempenho em intervalos intradiários, funcionando de maneira inversamente proporcional aos modelos mais complexos. Além disso, os autores destacam que o baixo volume de dados em cortes diários pode impactar negativamente a modelagem, uma vez que esses modelos não conseguem capturar os padrões de longo prazo de forma adequada.

Figura 16 - Processo de sumarização dos dados



Fonte: Chacón *et al.* (2023).

Em paralelo com os autores Rausch e Albrecht (2020) e Chacón *et al.* (2023), o trabalho de Albrecht, Rausch e Derra (2020) propõe uma ampliação no escopo das modelagens, incorporando mais técnicas de aprendizado de máquina. Além de *Random Forest* e *Gradient Boosting*, foi incluído a *K-Nearest Neighbor* e *Support Vector Regression*. Para fins de comparação, os autores utilizaram três modelos clássicos, com ênfase maior na modelagem ARIMA.

Foi utilizado uma base com ligações sobre reclamações de uma varejista online, com dados de janeiro de 2016 a maio de 2019, totalizando 175 semanas aproximadamente. Foi utilizado MAE e RMSE como métricas de performance dos modelos para previsões semanais, com a repetição deste processo ao decorrer de um ano. O melhor modelo ajustado foi o *Random Forest* com RMSE igual a 11,75, seguidos pelo *Gradient Boosting* com *Dropout*[1] e *Support Vector Regression*, com RMSE igual a 12,94 e 13,23, respectivamente. Com exceção do *K-Nearest Neighbor*, que obteve o pior desempenho com RMSE 18,21, a modelagem ARIMA aparece com RMSE de 14,53.

Os autores também destacam a importância do tempo de processamento dos modelos, observando que as modelagens clássicas exigem menor capacidade computacional em comparação com as abordagens de aprendizado de máquina. Eles enfatizam, portanto, a necessidade de avaliar cuidadosamente o *trade-off* entre o desempenho preditivo e os custos operacionais envolvidos na implementação desses modelos em ambientes de produção.

[1] O *dropout* permite a cada iteração do treinamento algumas árvores são ignoradas de forma aleatória, fazendo com que haja a diminuição de *overfitting*.

3 METODOLOGIA

Com base no conhecimento teórico em estatística e aprendizado de máquina estudados, a metodologia proposta foca no desenvolvimento de um modelo híbrido que combina abordagens clássicas e modernas, buscando alcançar previsões mais precisas. Adota-se uma abordagem explicativa para identificar e comparar o desempenho dos modelos priorizados, esclarecendo como cada um contribui para a solução do problema proposto. Além disso, a aplicação prática utilizou uma base de dados reais de um *call center*, seguida de uma análise comparativa para avaliar os resultados obtidos entre os modelos selecionados.

3.1 OBTENÇÃO BASE DE DADOS DE UM CALL CENTER

A partir de Services (2023) foi obtido o conjunto de dados utilizado contém o registro de todas as chamadas de Solicitação de Atendimento ao Cidadão (CSR) recebidas na central de atendimento da Secretaria de Serviços Públicos (DPS) na cidade de Cincinnati em Ohio, EUA. A escolha da base de dados foi fundamentada em sua facilidade de acesso, sendo pública e gratuita, além de abranger um alto volume de informações no período de 2014 a 2023. Outro fator determinante foi a riqueza de dados adicionais sobre as chamadas, que não apenas enriquecem o presente estudo, mas também oferecem potencial para análises e pesquisas futuras.

As CSR oferecem aos residentes de Cincinnati a oportunidade de enviar solicitações de serviço para questões como coleta de móveis, grama alta e reparos de buracos. As principais informações sobre a base estão demostradas na tabela abaixo:

Tabela 2 - Informações base de dados

Ano	Contagem Linhas	Contagem Colunas
2014	32.595	23
2015	110.687	23
2016	108.417	23
2017	101.183	23
2018	104.301	23
2019	105.273	23
2020	96.453	23
2021	101.117	23
2022	81.070	23
2023	295	23
Total	**841.391**	

Fonte: Elaborado pelo autor (2024).

O conjunto de dados está sumarizado por registros de cada ligação, contendo colunas com informações como data de início e fim, status da chamada, status de abandono, entre outros, podendo ser visualizado na Figura 17.

Figura 17 - Conjunto de dados inicial

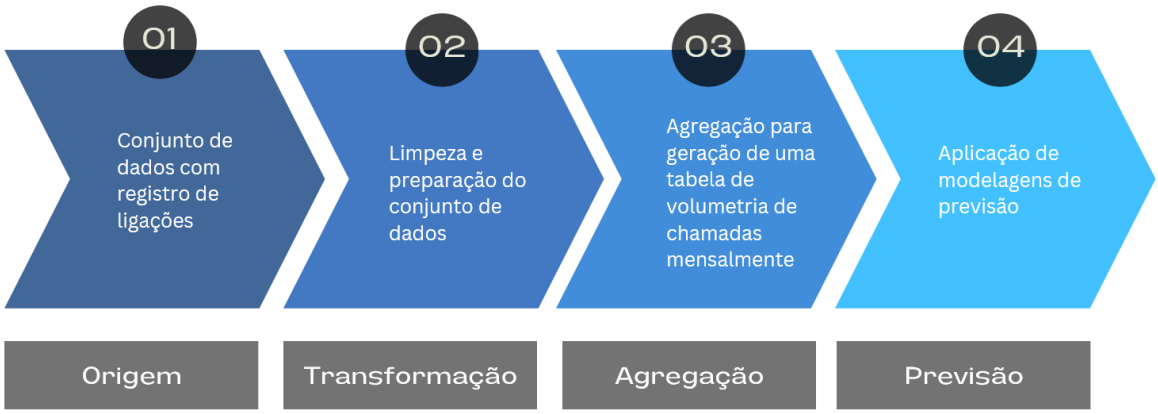
registroId	dtChamada	acaoChamadaId	acaoMotivoChamadaId	chamadaId	....	tempoRespostaSeg	tempoConversaSeg	tempoEncerramentoSeg	nivelServico	statusAbandono	statusChamada
93093223	12/09/2014	3	2	3	...	0	238	0	0	0	1
1,173E+10	23/06/2015	3	2	492	...	0	82	0	0	0	1
93083687	12/09/2014	3	2	7	...	0	125	0	0	0	1
2,895E+11	15/11/2019	21	4	4486	...	23	0	0	1	0	0
.	.	.	.	.		.	.	.	.	.	.
.	.	.	.	.	⋱	.	.	.	.	.	.
.	.	.	.	.		.	.	.	.	.	.

5 rows × 23 columns

Fonte: Elaborado pelo autor (2024).

Para o escopo do projeto, será utilizada uma tabela agrupada que apresenta a volumetria de chamadas por mês que será transformada conforme fluxo da Figura 18. Toda a transformação e visualização dos dados será detalhada na seção dedicada aos resultados.

Figura 18 - Fluxo de estruturação conjunto de dados

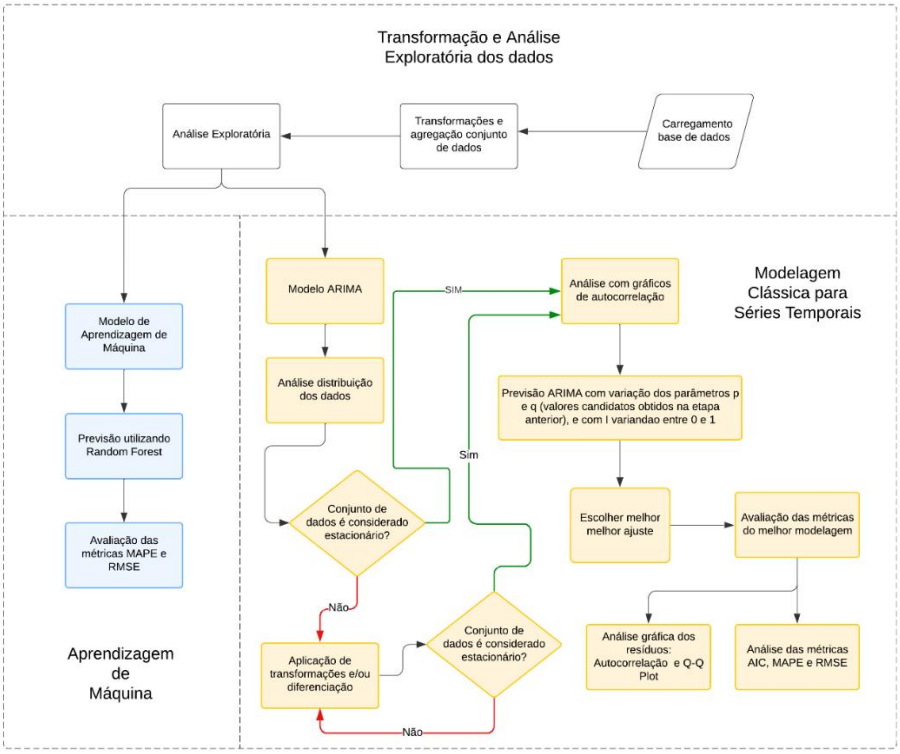


Fonte: Elaborado pelo autor (2024).

3.2 TRATAMENTO E LIMPEZA DOS DADOS COLETADOS

Para melhor entendimento da abordagem utilizada, foi representado na Figura 19 todo o fluxo adotado desde a obtenção da base até a etapa de previsão.

Figura 19 - Fluxograma da abordagem utilizada no trabalho



Fonte: Elaborado pelo autor (2024).

Para desenvolvimento das etapas propostas ilustrado na Figura 19, foi fundamental escolher um ferramental que proporcionasse flexibilidade e desempenho para aplicabilidade de todas as técnicas para transformações e manipulação dos dados, visualizações gráficas e utilização dos modelos.

Identificou-se diversas opções para a análise de dados, incluindo linguagens de programação como Python e R, além de softwares especializados como MINITAB, SPSS e MATLAB. Contudo, com base em experiências profissionais prévias, utilizou a linguagem de programação Python, que oferece uma ampla disponibilidade de bibliotecas que não apenas simplificam a obtenção e manipulação de dados, mas também facilitam a aplicação de modelos preditivos de forma eficiente e flexível.

Segundo Ramalho (2015 apud Python Tutorial Oficial, p. 17) “[...] Python é uma linguagem de programação poderosa e fácil de aprender [...]”.

A linguagem Python para muitos é denominado um framework, principalmente pela facilidade de implementar métodos e funções, ampla gama de funcionalidades, como manipulações de datas, leitura de diversos tipos de arquivos e a possibilidade de instalações de bibliotecas completas que executam uma atividade específica. Além disso, todos os principais pacotes utilizados são mantidos pela própria comunidade, devido ser um programa de código aberto (*open source*). (Ramalho, 2015)

Para o desenvolvimento, foi utilizado o *software Microsoft Visual Code*, que é um ambiente de desenvolvimento que possibilita a codificação e compilação em linguagem Python. Foi utilizada a versão 3.12.4 do Python e as principais bibliotecas foram:

- *pandas:* Utilização para manipulação e tratamento dos dados;
- *numpy:* Utilização para operações matemáticas;
- *matplotlib:* Utilização para geração de visualizações gráficas;
- *statsmodels:* Utilização para análise e previsão com ARIMA;
- *sklearn:* Utilização para previsão com *Random Forest* e avaliação das métricas de erros das modelagens.

3.3 ANÁLISE EXPLORATÓRIA DOS DADOS

Como mencionado no início do capítulo, para aplicar as modelagens propostas, é necessário realizar algumas transformações na base de dados e análises iniciais.

Nesta etapa, utilizando os registros de ligações referentes ao período de abril de 2014 a

outubro de 2023, foram iniciadas as transformações do conjunto de dados e a análise exploratória, com foco principal no tratamento de outliers. A base foi disponibilizada em formato CSV (arquivo separado por vírgulas) e importada usando a linguagem Python e a biblioteca pandas.

Após essa transformação, iniciou-se a análise exploratória para entender melhor a distribuição dos dados, visualizar a volumetria resumida e identificar possíveis outliers. As análises realizadas serão detalhadas nos resultados.

3.4 PRIORIZAÇÃO DE MODELOS

A decorrer do trabalho foi possível constatar uma predominância no uso do modelo ARIMA para modelagens clássicas de séries temporais. Além disso, o Random Forest tem sido amplamente aplicado no contexto utilizando AM, sendo eficaz para aplicações envolvendo séries temporais. Diante dessa constatação, o foco de desenvolvimento vai ser estudar os dois métodos aplicado a uma base real.

3.4.1 Modelagem Clássica para Séries Temporais (ARIMA)

A aplicação utilizando ARIMA seguiu os passos abaixo:

1. Análise distribuição dos dados: a partir dos gráficos de histograma e Q-Q Plot, verificar visualmente se o conjunto de dados apresenta características de estacionaridade a partir da visualização da aproximação de uma distribuição gaussiana (normal).
2. Aplicação transformações e/ou diferenciação: caso os dados não sejam considerados estacionários, aplicar transformações como logaritmo ou diferenciação para alcançar estacionaridade.
3. Especificação - Análise com gráficos de autocorrelação: com o conjunto de dados agora estacionário, utilizar os gráficos de autocorrelação (FAC) e autocorrelação parcial (FACP) para identificar os melhores candidatos para as ordens p e q da modelagem ARIMA.
4. Identificação - Previsão utilizando ARIMA: Dividir os dados em conjunto de treino e teste (80% para treino e 20% para teste). Realizar ajustes no modelo ARIMA com base nos valores observados nos gráficos de autocorrelação.
5. Diagnóstico - Seleção melhor ajuste e avaliação do modelo: Escolher o melhor ajuste do modelo ARIMA com base nas métricas de AIC, MAPE e RMSE, análise dos resíduos

e visualização dos resíduos em gráficos de autocorrelação e Q-Q Plot.

3.4.2 Aprendizagem de Máquina (Random Forest)

Para aplicação da modelagem *Random Forest Regressor* foi realizado os seguintes passos:

1. Preparação dos dados: criação de uma coluna com o valor defasado (*lag*) de um mês para capturar a dependência temporal;
2. Divisão dos dados: divisão do conjunto de dados em dados de treino e teste (no caso, foi 80% para treino e 20% para teste);
3. Treinamento *Random Forest*: configuração o modelo de Random Forest para regressão, especificando e testando parâmetros como a quantidade de árvores utilizado na modelagem (*n_estimators*), e aleatoriedade da previsão (*random_state*).
4. Realização de previsões no conjunto teste: Gerar previsões no conjunto de teste e armazená-las para análise comparativa.
5. Visualização das previsões: Plotar as séries de treino, teste e previsões para avaliar visualmente o ajuste com relação aos dados reais.
6. Avaliação do desempenho do modelo: Calcular métricas de desempenho, como o MAPE e RMSE, para avaliar a precisão do modelo.

4 RESULTADOS

Neste capítulo 4 será abordado o detalhamento das análises do conjunto dado de um *call center* responsável por atender solicitações dos cidadãos da cidade de Cincinatti, no EUA. Para melhor entendimento, o tópico está estruturado a partir dos tópicos: 4.1 TRATAMENTO E LIMPEZA DOS DADOS COLETADOS, 4.2 ANÁLISE EXPLORATÓRIA DOS DADOS, 4.3 MODELAGEM ARIMA e 4.4 MODELAGEM RANDOM FOREST.

4.1 TRATAMENTO E LIMPEZA DOS DADOS COLETADOS

4.1.1 Tratamento de Outliers

O conjunto completo de dados importado continha 841.391 linhas e 23 colunas conforme foi ilustrado no item 3.1, sendo transformado em uma tabela sumarizada e ordenada (ordem crescente) com duas colunas: mesAno e qtdChamada, onde mesAno representa o agrupamento de mês e ano, e qtdChamada a quantidade de registros correspondentes em cada mês. Primeiramente, verificou-se havia algum mês faltante, foi contabilizado um total esperado de 115 meses no intervalo de 04/2014 a 10/2023. Confirmou-se que todos os meses estavam presentes. A Tabela 3 é possível visualizar uma amostra do conjunto de dados resultante dessa análise.

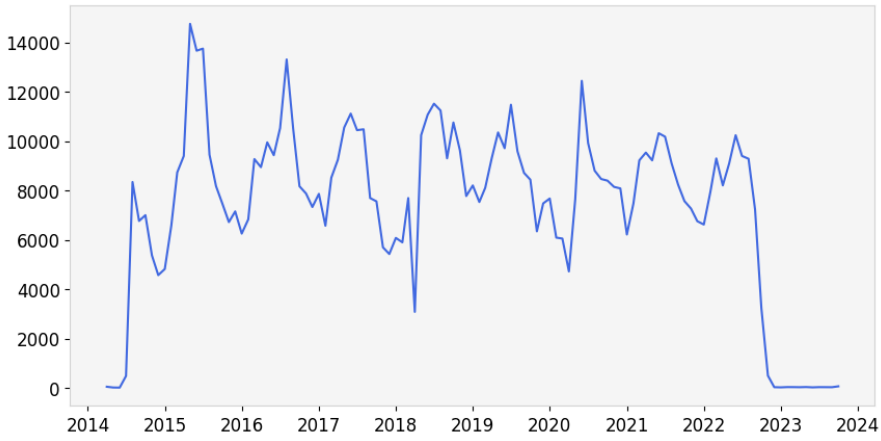
Tabela 3 - Amostragem do conjunto de dados agrupado

indice	mesAno	qtdChamada
1	01/04/2014	44
2	01/05/2014	13
3	01/06/2014	9
4	01/07/2014	480
5	01/08/2014	8345
⋮	⋮	⋮
111	01/06/2023	19
112	01/07/2023	28
113	01/08/2023	27
114	01/09/2023	25
115	01/10/2023	61

Fonte: Elaborado pelo autor (2024).

Pode-se notar que valores dos primeiros e últimos meses do conjunto de dados são valores baixos, e aparentemente estão destoando dos demais, isso pode ser mais bem visualizado na Figura 20. Através do gráfico verifica-se que há duas regiões com valores extremamente menores que os demais, no início e fim, que possivelmente são outliers.

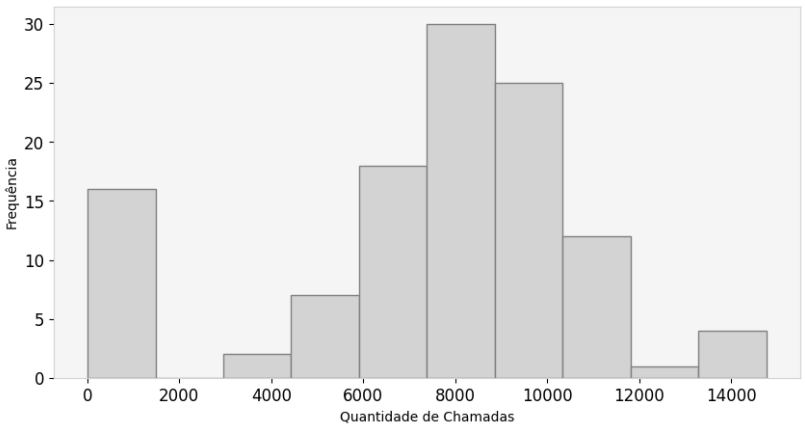
Figura 20 - Quantidade de chamadas ao decorrer dos meses



Fonte: Elaborado pelo autor (2024).

Para analisar mais precisamente a presença de valores atípicos utilizou os gráficos histograma e *boxplot* para entender a distribuição das volumetrias. Inicialmente, foi verificado via histograma que os dados tendem a se aproximar de uma distribuição normal. No entanto, observou-se uma grande concentração de valores entre 0 e 2000, e outro grupo menor com concentração atípica acima de 13000. Esses valores destoam dos padrões gerais da base de dados, conforme ilustrado na Figura 21.

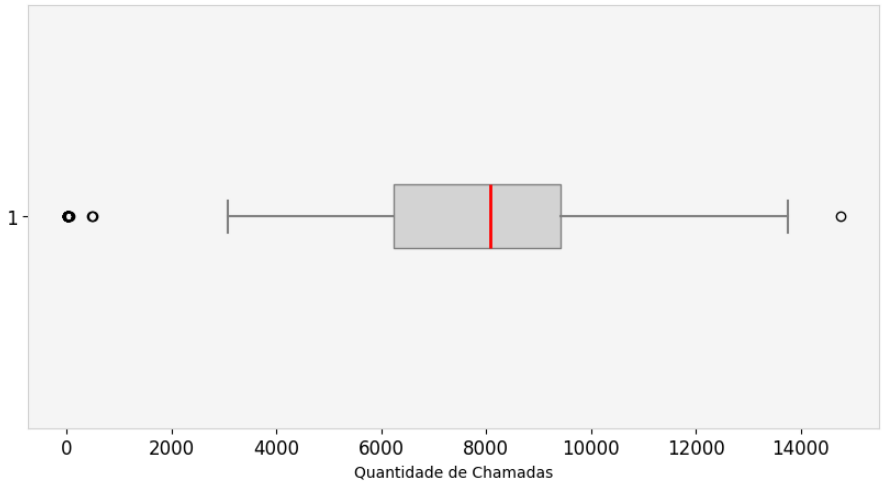
Figura 21 - Histograma de quantidade de chamadas



Fonte: Elaborado pelo autor (2024).

Para identificação dos outliers mais especificamente, foi utilizado o b*oxplot* ilustrado na Figura 22, em que os pontos além dos limites da caixa são valores atípicos, e com isso foi possível determinar os limites inferiores e superiores.

Figura 22 - *Boxplot* de quantidade de chamadas



Fonte: Elaborado pelo autor (2024).

As volumetrias consideradas outliers estão representados na Tabela 4, a partir dessa informação foi possível confirmar que a maioria se trata de valores iniciais ou finais do conjunto de dados, com exceção do registro 14, do mês de maio de 2015. Para prosseguir com as previsões, decidiu-se por excluir os dados considerados outliers, exceto o registro 14. A baixa quantidade de ligações nesses meses pode ter diversas causas, mas, devido ao afastamento tão significativo em relação à mediana histórica, é provável que tenha ocorrido algum problema na ingestão dos dados. Esses dados excluídos, portanto, não interferirão nas previsões futuras.

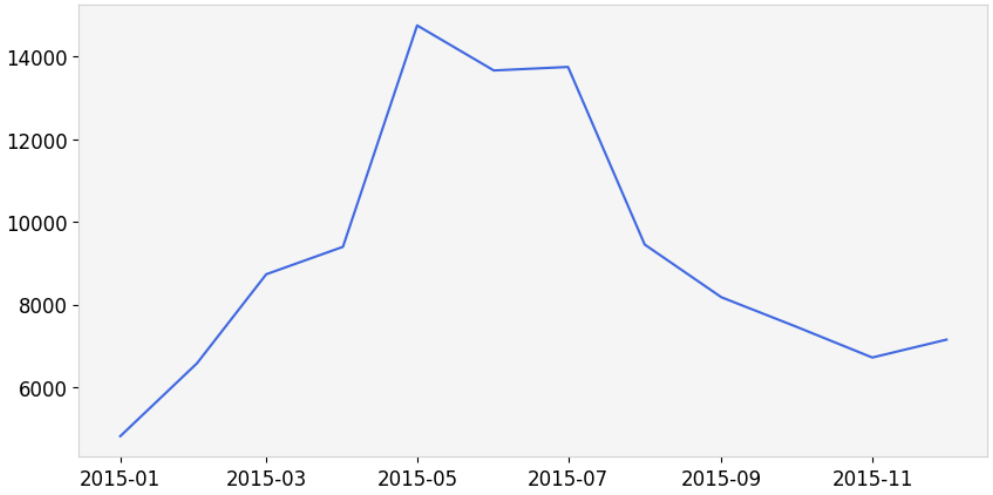
Tabela 4 - Meses considerado outliers

indice	mesAno	qtdChamada
1	01/04/2014	44
2	01/05/2014	13
3	01/06/2014	9
4	01/07/2014	480
14	01/05/2015	14758
104	01/11/2022	493
105	01/12/2022	26
106	01/01/2023	19
107	01/02/2023	30
108	01/03/2023	29
109	01/04/2023	25
110	01/05/2023	32
111	01/06/2023	19
112	01/07/2023	28
113	01/08/2023	27
114	01/09/2023	25
115	01/10/2023	61

Fonte: Elaborado pelo autor (2024).

No entanto, referente ao registro 14 observa na Figura 23, que 2015 foi um ano com maiores volumetrias em picos sazonais comparados com mesmos períodos dos demais anos, acredita-se ser algo que pode acontecer a depender de variáveis externas, como eventos, economia e entre outros fatores. Neste caso, optamos por não excluir os dados, pois, embora sejam eventos atípicos, eles não prejudicarão significativamente as previsões. A exclusão desses dados poderia causar mais impacto, já que teríamos de desconsiderar toda a série anterior a 05/2015. Em séries temporais, a ausência de um mês teria uma influência considerável nas previsões.

Figura 23 - Visualização quantidade de chamadas em 2015



Fonte: Elaborado pelo autor (2024).

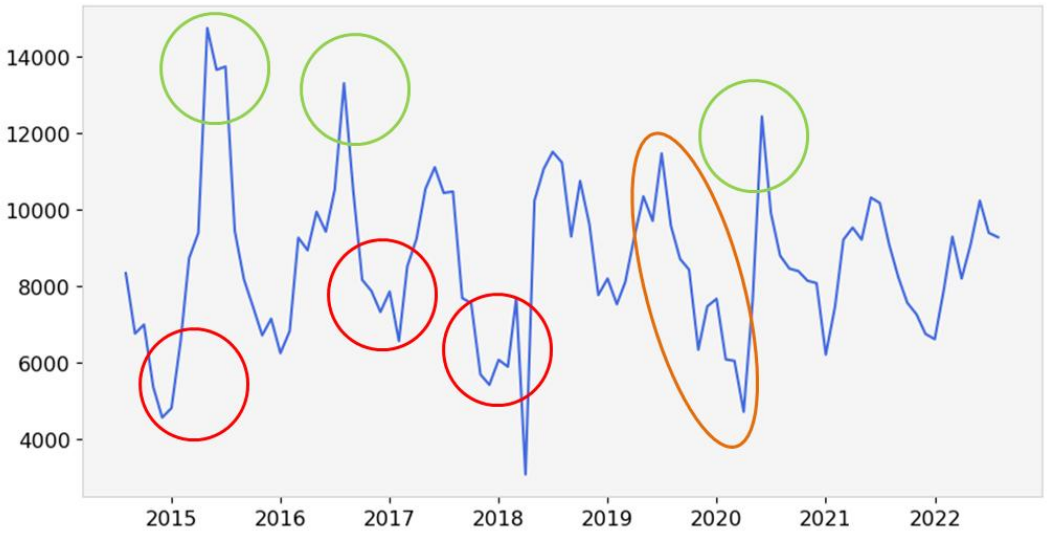
Por fim, antes de iniciar as análises, decidiu-se considerar o período de 08/2014 a 08/2022, desconsiderando os dois últimos meses de 2022. Assim, obtemos um intervalo simétrico de 8 anos completos.

4.2 ANÁLISE EXPLORATÓRIA DOS DADOS

4.2.1 Análise dos padrões da série

Com o conjunto de dados sem outliers, inicia-se a fase de exploração dos dados. Inicialmente no estudo foi realizado algumas análises gráficas: série história com a volumetria de chamadas as decorrer dos meses, histograma para visualizar a distribuição dos dados livre de outliers e *Q-Q Plot* para validar visualmente se os dados podem considerado uma distribuição normal. Na Figura 24 foi ilustrado a quantidade de chamadas ao decorrer dos meses, sendo desconsiderado os outliers.

Figura 24 - Quantidade de chamadas ao decorrer dos meses sem outliers



Fonte: Elaborado pelo autor (2024).

A partir da visualização da volumetria história é possível observar que a central de atendimento recebe um maior volume de ligações na metade do ano, e menores no começo e início do ano, conforme marcações em círculos verdes e vermelhos, respectivamente, na Figura 24.

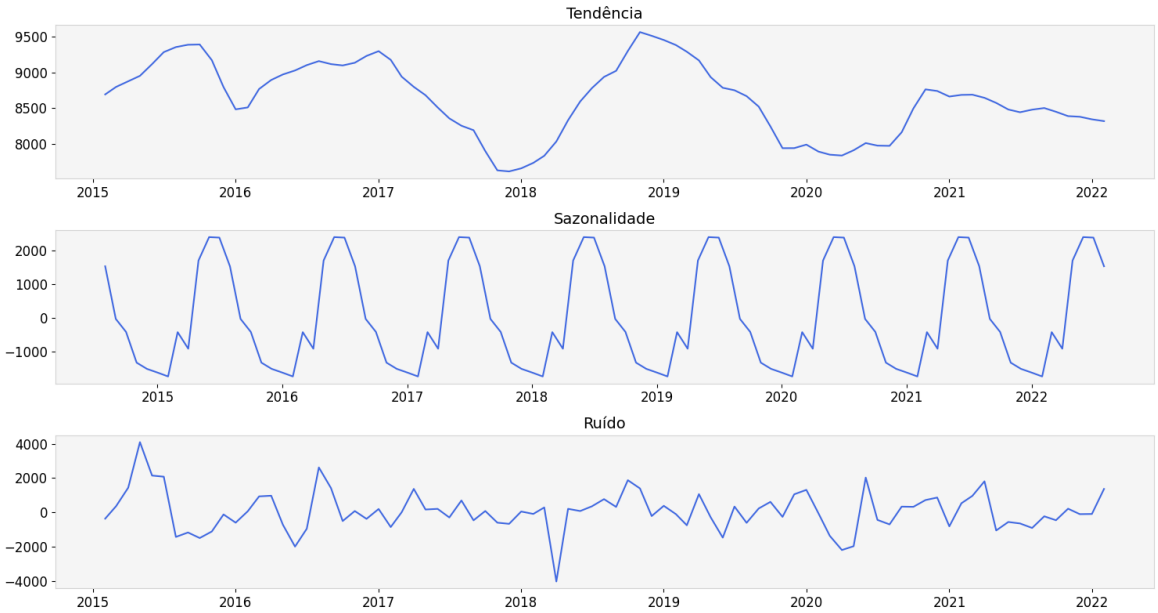
Considerando que a base de dados se refere a solicitações de serviços públicos, como coleta de móveis, corte de grama e reparos em vias, é possível associar essas demandas à presença das pessoas nas ruas ou ao aumento das atividades de manutenção e construção. Assim, podemos relacionar esses padrões com as estações do ano. Como a cidade de Cincinnati está localizada no hemisfério norte, o verão ocorre no meio do ano, época em que há um maior fluxo de pessoas nas ruas e mais atividades de construção e reforma, elevando a demanda por serviços públicos. Já no final e início do ano, o inverno inibe essas atividades, o que resulta em uma dinâmica oposta à do verão.

A partir da Figura 25 revela que a quantidade de chamadas apresenta uma tendência irregular, sem um padrão consistente. Observa-se uma queda acentuada no volume de chamadas ao final de 2019, que se estende durante todo o ano de 2020, com uma recuperação apenas parcial até 2022, podendo ser visualizado pela marcação em laranja na Figura 25. Essa diminuição coincide com o início e o auge da pandemia de COVID-19, que em 2020 levou a medidas restritivas que limitaram a mobilidade e modificaram drasticamente o dia a dia das pessoas. Assim, é provável que as restrições e o isolamento social, necessários para conter o vírus, tenham impactado o número de chamadas, já que muitos serviços e fluxos de trabalho

presenciais foram interrompidos ou adaptados para o modelo remoto.

Além disso é possível verificar picos de ruídos em 2015 e 2018, que podem indicar eventos fora do padrão, possivelmente influenciados por fatores externos específicos.

Figura 25 - Decomposição da série em tendência, sazonalidade e ruído



Fonte: Elaborado pelo autor (2024).

A Figura 25 revela também de maneira mais clara a sazonalidade discutida anteriormente, onde é possível verificar que os picos de demandas em meados dos anos e quedas acentuadas no início e fim deles. Na análise do gráfico de ruídos, é evidente que há flutuações aleatórias atípicas em alguns períodos. Observa-se, por exemplo, um pico positivo em meados de 2015 e um pico negativo de valor absoluto similar no início de 2018. Esse comportamento de ruído deveria idealmente apresentar uma média tendendo a zero e uma variância constante ao longo do tempo, indicando variações aleatórias ao redor do ponto médio estável.

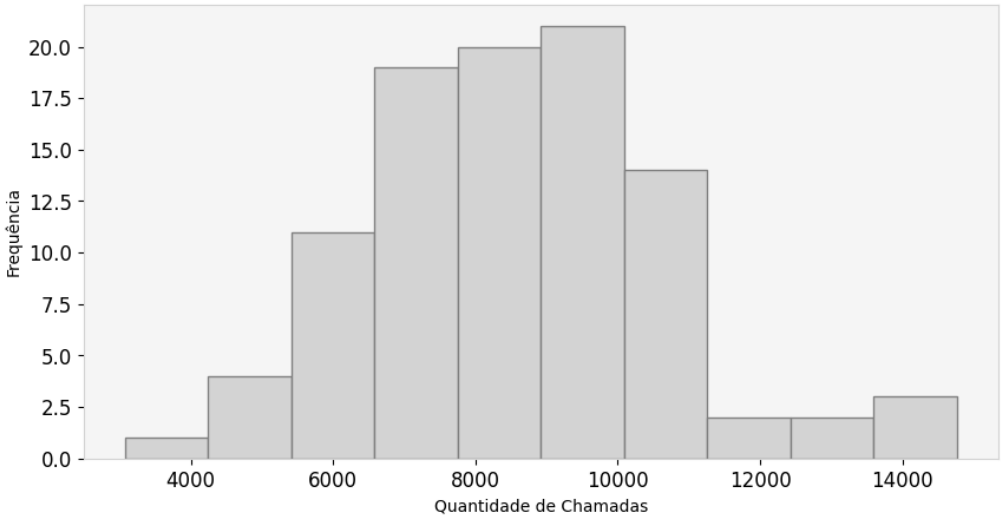
Essas flutuações sugerem a existência de eventos fora do padrão, que podem ter sido influenciados por fatores externos ou mudanças no contexto do serviço público, que contribuíram para desviaram os valores.

4.2.2 Análise distribuição dos dados

Após a remoção de alguns meses, por apresentarem volumetrias consideradas outliers

de acordo com análises anteriores, obtém-se um gráfico de histograma mais uniforme, ilustrado na Figura 26.

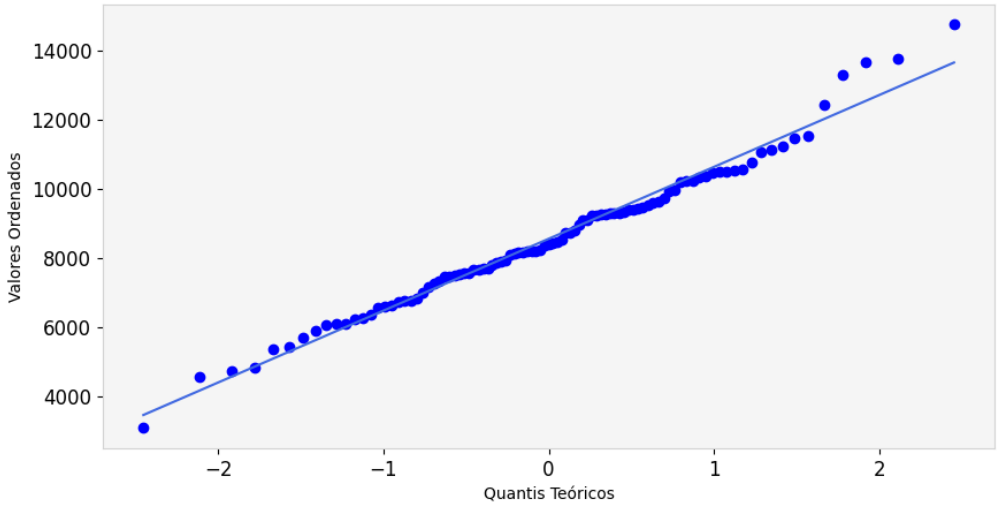
Figura 26 - Histograma de quantidade de chamadas sem outliers



Fonte: Elaborado pelo autor (2024).

Além do histograma foi realizado a plotagem do gráfico *Q-Q Plot*, ilustrado na Figura 27. Com base nos gráficos analisados, pondera-se que o conjunto de dados tende a uma distribuição normal. Baseando-se tanto no formato de sino observado no histograma, característico da curva normal, quanto na correlação linear entre os intervalos dos dados observados e os intervalos teóricos de uma distribuição normal.

Figura 27 - *Q-Q Plot* dos dados



Fonte: Elaborado pelo autor (2024).

Os aspectos gerados sugerem que os dados seguem, em sua maioria, uma distribuição normal, o que permite assumir que o conjunto de dados pode ser considerado estacionário. Com essa inferência de estacionariedade, não houve necessidade de transformações como a de aplicar logaritmo ou a diferenciação, que contribuem para estabilizar a média e a variância dos dados ao longo do tempo.

4.3 MODELAGEM ARIMA

Conforme já mencionado em metodologia, a modelagem ARIMA para previsão segue cinco etapas principais. O primeiro passo, já descrito anteriormente, envolve a verificação da estacionariedade da série de dados. No segundo passo, geralmente é necessário realizar transformações (como logaritmos ou diferenciação) para estabilizar as séries que não a apresentam inicialmente. No entanto, o conjunto de dados trabalhado apresentou um comportamento estacionário desde o início, conforme ponderações já descritas.

Neste tópico, trataremos as demais etapas: Especificação - Análise com gráficos de autocorrelação; Identificação - Previsão utilizando ARIMA; Diagnóstico - Seleção melhor ajuste e avaliação do modelo.

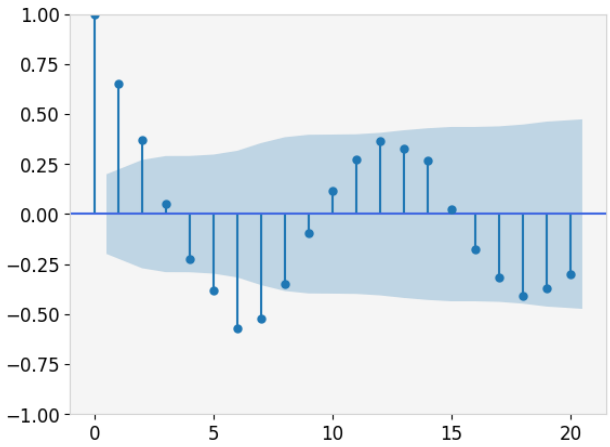
4.3.1 Especificação - Análise com gráficos de autocorrelação

Para determinar os possíveis valores de p e q, utilizaremos os gráficos de Função Autocorrelação (FAC) e Função Autocorrelação Parcial (FACP), ilustrados na Figura 28 e Figura 29. Observa-se que o gráfico FAC, há um decaimento exponencial com cortes claros em certos *lags* que ultrapassam a região de 95% de significância, indicando a influência da estrutura de Médias Móveis, MA(q). Os *lags* significativos aparecem em 1, 2, 5, 6 e 7, o que indica que o parâmetro q pode ser um desses valores, com maior indicativo nos valores extremos, como 1 e 6, que apresentam maiores probabilidades.

Esses *lags* significativos sugerem que a inclusão de componentes de médias móveis na modelagem pode ser capaz de reproduzir os padrões encontrados na série estudada e capturar melhor as flutuações aleatórias e os desvios ao longo do tempo, otimizando a previsão e estabilizando a série em torno de seu valor esperado. Esses gráficos foram cruciais para identificação dos padrões de dependência temporal nos dados, auxiliando no levantamento dos

parâmetros ideais para o modelo ARIMA.

Figura 28 - Gráfico de autocorrelação de chamadas

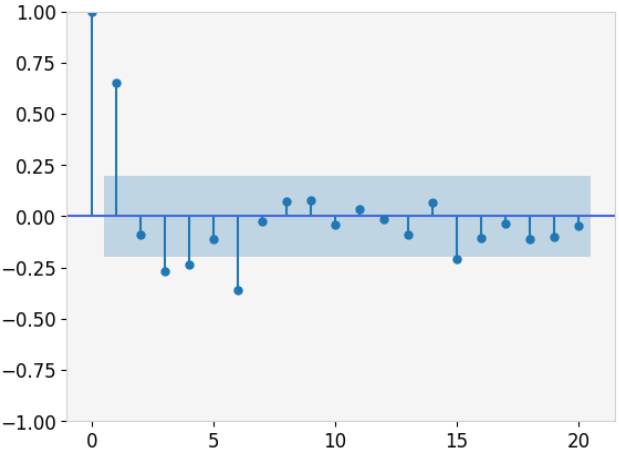


Fonte: Elaborado pelo autor (2024).

Observa-se que o gráfico FAC, há um decaimento exponencial com cortes claros em certos *lags* que ultrapassam a região de 95% de significância, indicando a influência da estrutura de Médias Móveis, MA(q). Os *lags* significativos aparecem em 1, 2, 5, 6 e 7, o que indica que o parâmetro *q* pode ser um desses valores, com maior indicativo nos valores extremos, como 1 e 6, que apresentam maiores probabilidades.

Esses *lags* significativos sugerem que a inclusão de componentes de médias móveis na modelagem pode ser capaz de reproduzir os padrões encontrados na série estudada e capturar melhor as flutuações aleatórias e os desvios ao longo do tempo, otimizando a previsão e estabilizando a série em torno de seu valor esperado.

Figura 29 - Gráfico de autocorrelação parcial de chamadas



Fonte: Elaborado pelo autor (2024).

O gráfico FACP apresenta um comportamento esperado para uma série, com um decaimento exponencial e cortes bem definidos. Isso sugere a existência do termo Autoregressivo, AR(p), com *lags* significativos em 1, 3, 4, 6 e 15. Entre esses, os *lags* 1 e 6 parecem ter maior potencial para definir o valor de p, indicando uma influência de valores próximos e, em menor grau, de valores mais distantes (como o *lag* 6).

Quanto ao parâmetro de diferenciação (d), foi utilizado o valor 0 (pois a série foi considerada estacionária), mas também, d igual a 1, para examinar se há uma influência direta dos valores antecessores nos posteriores a ele. Esse teste permitirá definir o melhor ajuste de modelo ARIMA, ajudando a refinar a escolha dos parâmetros e melhorar a precisão da previsão.

Sabendo disso, a próxima etapa foi avaliar os ajustes com as possíveis variações de p, q e d para identificar o modelo que melhor representa a série e reduz erros de previsão.

4.3.2 Identificação - Previsão utilizando ARIMA

Primeiramente, a base de dados foi dividida entre conjuntos de treino e teste, utilizando 80% dos dados para treino e os 20% restantes para teste. Inicialmente, a base foi ordenada pela coluna “mesAno”, da data mais antiga para a mais recente. Dessa forma, o intervalo de agosto de 2014 a dezembro de 2020 foi selecionado para o conjunto de treinamento, enquanto o período de janeiro de 2021 a agosto de 2022 foi reservado para o teste.

A partir da lista de parâmetros (p, q e d), foi modelado utilizando o conjunto de treino e realizado a previsão utilizando a base teste. Obteve-se um plural de variações dos parâmetros e para escolha das modelagens representada na Tabela 5, no qual é exibido os 5 melhores ajustes de cada métrica.

Tabela 5 - Melhores modelagens obtidas

Modelo	AIC	MAPE	RMSE	Classificação por AIC	Classificação por MAPE	Classificação por RMSE
ARIMA(4,0,6)	**1350,1**	**12,3%**	**1251,7**	**1**		
ARIMA(4,0,5)	1350,5	11,4%	1169,3	2		
ARIMA(4,1,6)	1351,1	9,8%	1011,7	3	3	5
ARIMA(4,0,7)	1351,9	12,9%	1266,2	4		
ARIMA(6,0,5)	1351,9	12,1%	1232,3	5		
ARIMA(6,1,6)	**1354,3**	**9,5%**	**973,5**		**1**	**2**
ARIMA(6,1,7)	1358,9	9,8%	1005,0		2	4
ARIMA(6,1,5)	1353,6	10,3%	1073,3		4	

ARIMA(6,0,2)	**1358,8**	**10,4%**	**972,6**	**5**	**1**
ARIMA(6,0,1)	1356,9	10,7%	1001,0		3

Fonte: Elaborado pelo autor (2024).

Dentre os modelos analisados, o ARIMA(4,0,6) se destaca por apresentar o melhor valor de AIC, indicando um equilíbrio ideal entre ajuste e complexidade do modelo, o que contribui para a redução do risco de *overfitting*. Em relação à métrica MAPE, o modelo ARIMA(6,1,6) se mostra como o melhor ajuste, apresentando o menor erro absoluto entre os valores previstos e reais. Além disso, esse modelo ocupa a segunda posição quando avaliado pelo RMSE, que representa o erro na quantidade real de chamadas. Por outro lado, o modelo ARIMA(6,0,2) exibe o melhor valor de RMSE, embora a diferença em relação ao segundo colocado seja bastante sutil.

Dentre os ajustes Tabela 5, optou em escolher 2 modelagens para analisar graficamente, são eles:

1. ARIMA (6,1,6): ótimo desempenho considerando as métricas MAPE e RMSE, mesmo que o valor de AIC não esteja entre os primeiros, a variação dele comparado com o primeiro é de apenas 0,64%;
2. ARIMA (4,0,6): apesar das diferenças de 29,47% e 28,70% nas métricas MAPE e RMSE em relação aos melhores modelos em cada uma, o foco aqui foi priorizar o desempenho na métrica AIC. A escolha de um modelo com menor AIC indica uma preocupação maior com o ajuste geral dos dados à modelagem, o que é particularmente relevante para previsões futuras. Essa abordagem favorece a seleção de um modelo robusto que, embora possa não ser o mais preciso em termos de erro percentual ou quadrático, oferece uma boa capacidade de adaptação e generalização para dados ainda não observados.

O trade-off entre AIC, MAPE e RMSE é crucial para uma modelagem eficaz, e a intenção foi realizar um balanço cuidadoso entre essas métricas. O modelo ARIMA(6,1,6) M1 se destaca nas métricas MAPE e RMSE, indicando sua capacidade de fornecer previsões mais precisas e confiáveis. Por outro lado, o ARIMA(4,0,6) M2 apresenta um ótimo desempenho em relação ao AIC, com uma performance abaixo para MAPE e RMSE.

4.3.3 Diagnóstico - Seleção melhor ajuste e avaliação do modelo

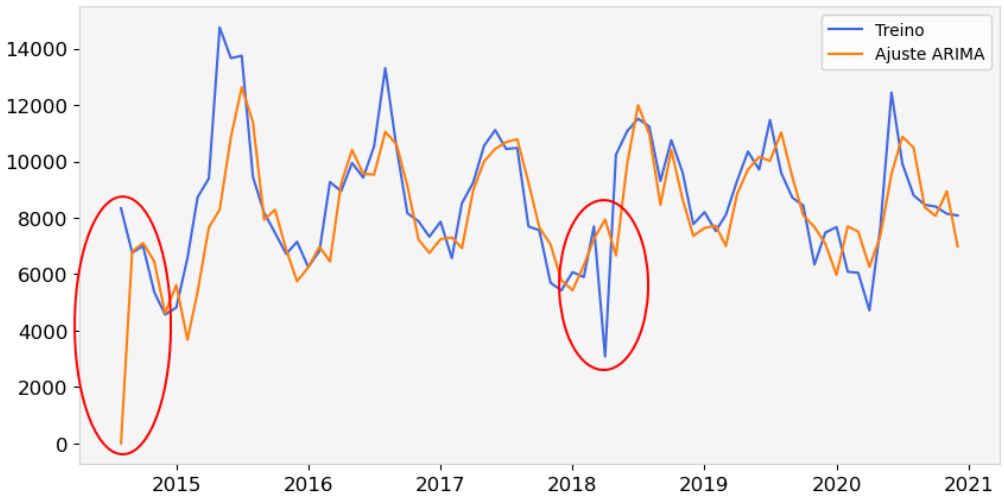
Como abordado em capítulos anteriores, em modelos de séries temporais utilizando

ARIMA, é fundamental não apenas considerar as métricas numéricas, mas também realizar uma avaliação gráfica do ajuste e das previsões. Gráficos como o *Q-Q Plot* e de autocorrelação dos resíduos são ferramentas essenciais que auxiliam na escolha da modelagem que melhor se adapta aos dados. Esses elementos visuais proporcionam uma compreensão mais profunda da adequação do modelo, permitindo identificar padrões não capturados e validar a suposição de normalidade dos resíduos.

4.3.3.1 Modelagem ARIMA (6, 1, 6)

O ajuste da base de treinamento pode ser visualizado a partir da Figura 30, onde a curva do ajuste do modelo é representada em laranja, enquanto a linha real da base de treinamento é exibida em azul. Observa-se que o modelo apresenta um bom ajuste aos dados; no entanto, há dois pontos que destoam dos demais, especificamente, no início de 2014, onde o modelo previu valores inferiores à realidade, e, em contraste, no primeiro terço de 2018, onde a previsão foi superior. Os pontos destacados estão indicados na Figura 30 pelas demarcações em círculos vermelhos. Esses desvios podem estar relacionados ao fato de o AIC não ter sido priorizado neste modelo. A seguir, realizaremos uma análise do outro ajuste para uma compreensão mais aprofundada.

Figura 30 - Ajuste no conjunto de treino M1

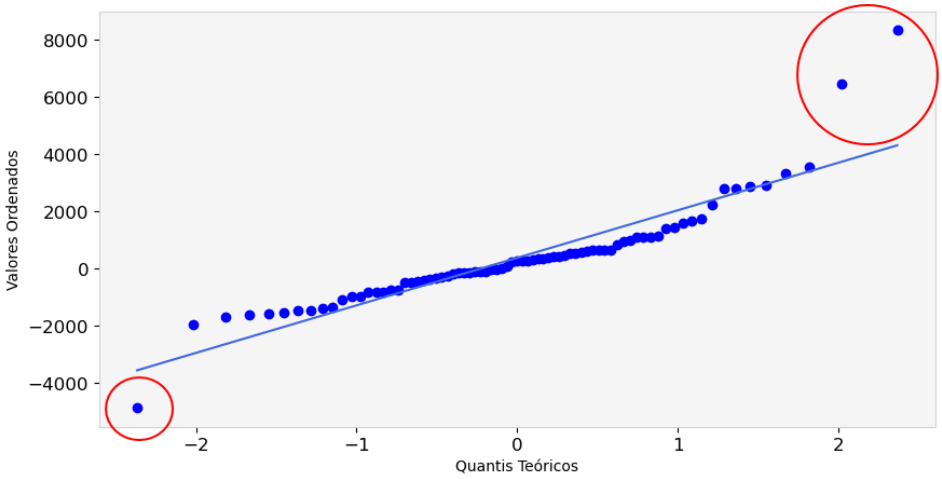


Fonte: Elaborado pelo autor (2024).

A partir dos gráficos *Q-Q Plot* ilustrado na Figura 31 é possível ponderar que os resíduos do modelo o conjunto em sua maioria se aproxima de uma distribuição normal, porém há três pontos fora da curva que indicam que há resíduos que se desviam das suposições de

normalidade. Esses pontos extremos podem sugerir que a modelagem pode não estar capturando completamente a variabilidade nos dados, possivelmente devido algum outliers que não conseguimos eliminar ou a uma estrutura não modelada adequadamente (como sazonalidade).

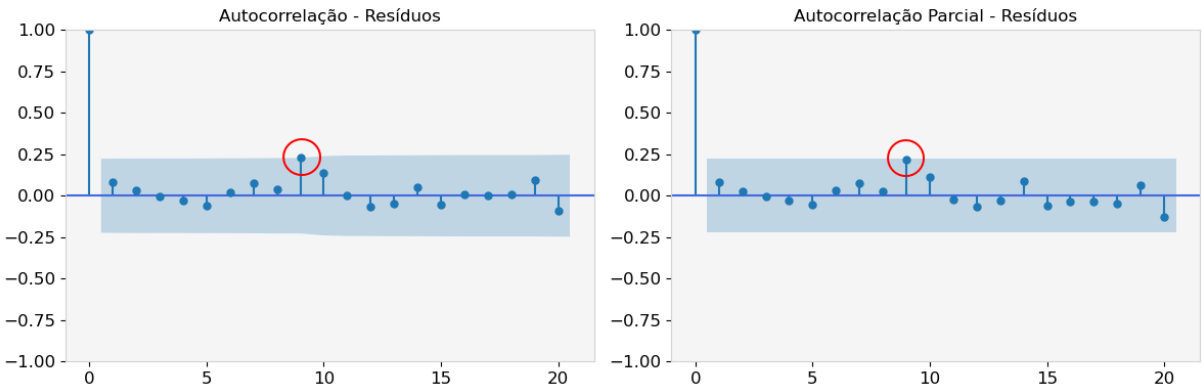
Figura 31 - *Q-Q Plot* para resíduos M1



Fonte: Elaborado pelo autor (2024).

Os gráficos de autocorrelação, ilustrados na Figura 32, mostram um comportamento quase ideal para os resíduos do modelo. Como esperado, todos os *lags* encontram-se dentro do intervalo de confiança estatística de 95%, exceto um ponto específico no *lag* 9, onde há uma leve saliência fora do intervalo de confiança em ambos os gráficos. Esse comportamento indica uma possível correlação residual neste *lag*, sugerindo que ainda há alguma estrutura não capturada pelo modelo, embora de forma sutil.

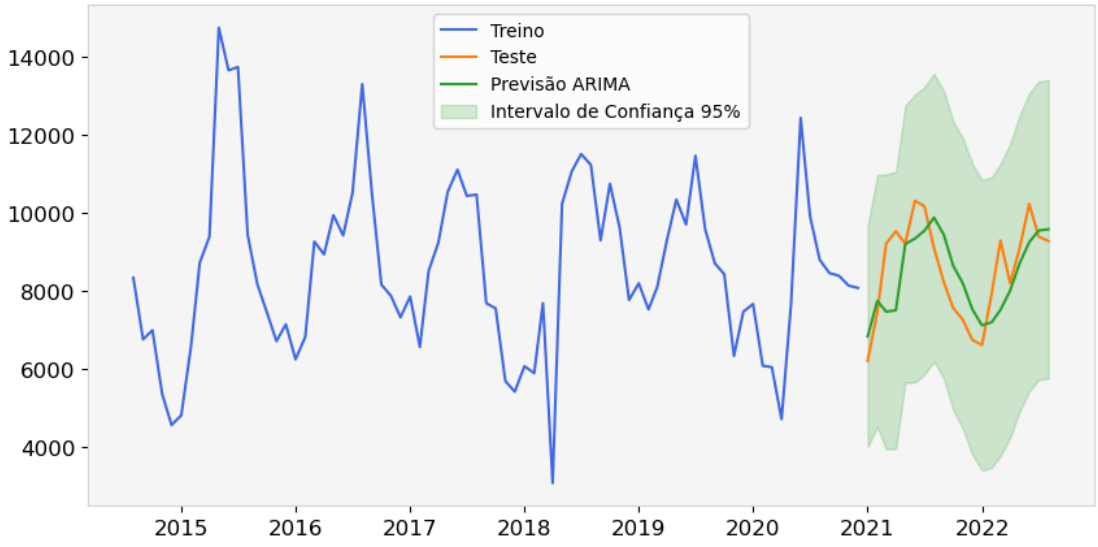
Figura 32 - Gráficos de autocorrelação para resíduos M1



Fonte: Elaborado pelo autor (2024).

Após as principais análises residuais, observa-se que, apesar de alguns pontos de atenção quanto à aderência dos dados ao ajuste, a Figura 33, indica um bom desempenho do modelo com essa configuração de parâmetros. Ao comparar as previsões com os dados reais da base de teste, verifica-se que as diferenças são pequenas, com apenas alguns pontos sutis que não foram completamente capturados pela modelagem. Isso demonstra uma precisão satisfatória nas previsões, mesmo considerando as pequenas variações não ajustadas.

Figura 33 - Gráfico comparativo entre previsto e realidade M1

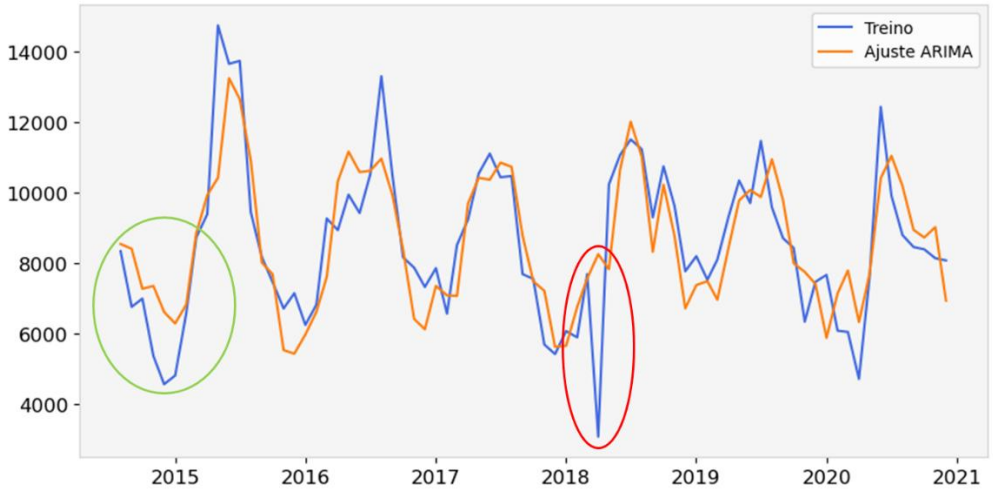


Fonte: Elaborado pelo autor (2024).

4.3.3.2 Modelagem ARIMA (4, 0, 6)

Na Figura 34, a configuração apresentou um desempenho superior no início de 2014, sendo que esse período foi identificado como um ponto insuficiente no modelo M1. Apesar desse desempenho inicial, o modelo ainda enfrenta dificuldades no primeiro terço de 2018, sugerindo que há uma limitação no seu poder preditivo para capturar padrões específicos ou mudanças estruturais neste período em específico.

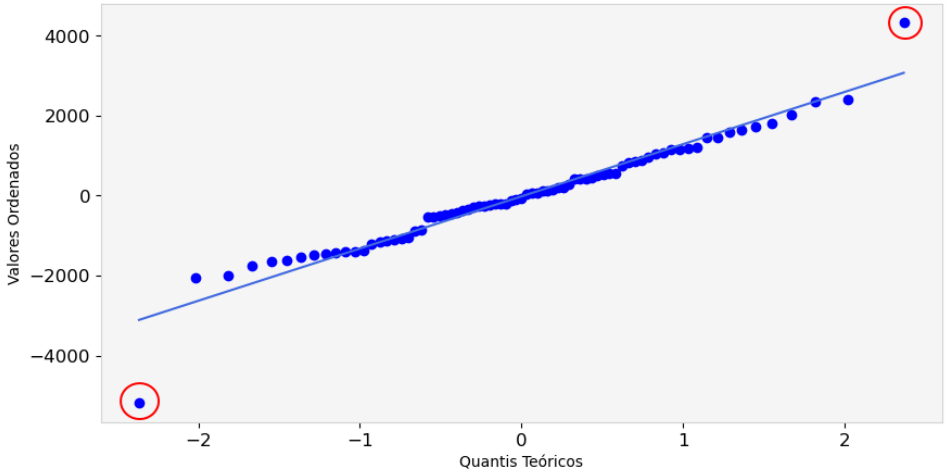
Figura 34 - Ajuste no conjunto de treino M2



Fonte: Elaborado pelo autor (2024).

No gráfico Q-Q Plot ilustrado na Figura 35, observa-se um bom alinhamento dos dados ao longo da linha de normalidade, o que sugere uma distribuição adequada dos resíduos em grande parte. No entanto, há dois pontos que se desviam dessa linha, indicando possíveis outliers ou áreas onde o modelo não capturou totalmente a variabilidade dos dados. Esses desvios são mínimos e, no contexto geral, não comprometem a qualidade do ajuste do modelo.

Figura 35 - *Q-Q Plot* para resíduos M2

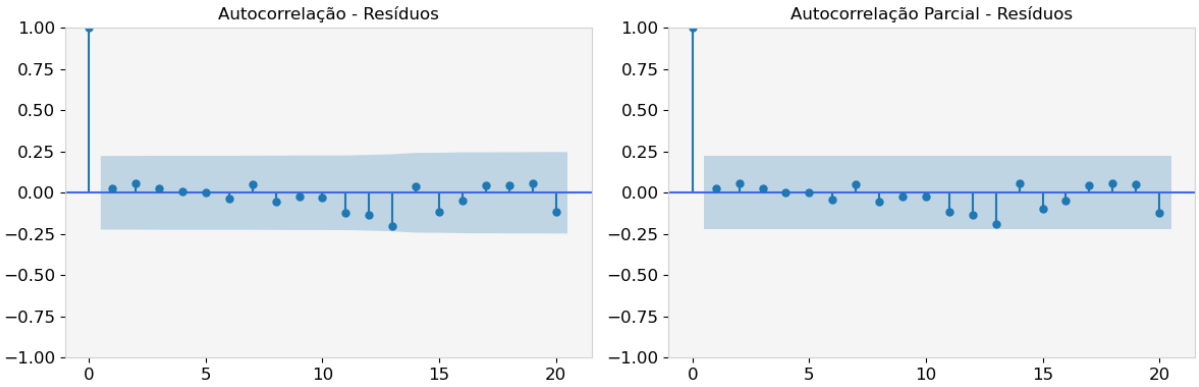


Fonte: Elaborado pelo autor (2024).

Para gráficos de autocorrelação ilustrado na Figura 36 exibem o comportamento esperado dos resíduos do modelo, com todos os *lags* permanecendo dentro do intervalo de

confiança estatístico de 95%. Esse resultado indica que não há correlação significativa entre os resíduos ao longo do tempo, sugerindo que o modelo capturou bem a estrutura dos dados e que os resíduos são, em grande parte, aleatórios. Essa condição é desejável para validar a suposição de que os resíduos não apresentam padrões não explicados pelo modelo, reforçando a adequação da modelagem.

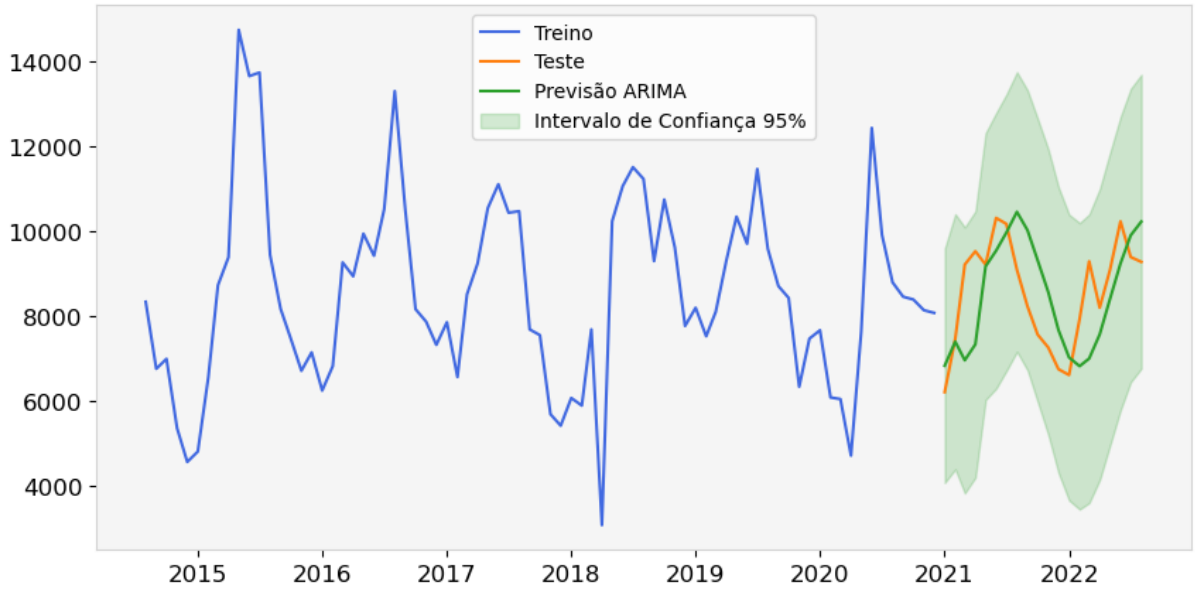
Figura 36 - Gráficos de autocorrelação para resíduos M2



Fonte: Elaborado pelo autor (2024).

Ao analisar a previsão no conjunto de teste e compará-la com os dados reais na Figura 37, nota-se uma leve melhora em relação ao ajuste do modelo M1, especialmente em meados de 2021. No entanto, visualmente, observa-se que, a partir desse período, as diferenças entre os valores reais e previstos aumentam, o que explica as métricas MAPE e RMSE com valores superiores em comparação ao modelo M1. Esse comportamento sugere que o modelo apresenta uma precisão aceitável no início do período de teste, mas sua capacidade de previsão se reduz conforme avançamos, refletindo as limitações de ajuste em cenários de maior variabilidade nos dados.

Figura 37 - Gráfico comparativo entre previsto e realidade M2



Fonte: Elaborado pelo autor (2024).

De maneira ampla, os modelos M1 e M2, avaliados com base nas análises gráficas, apresentam um bom ajuste aos dados. No entanto, é necessário considerar o trade-off constante entre a qualidade do ajuste e a capacidade de previsão da modelagem, refletido nas métricas AIC, MAPE e RMSE, respectivamente. Enquanto o AIC privilegia a capacidade de simplificar o modelo e evita o *overfitting*, o MAPE e o RMSE focam na precisão das previsões, cada um com uma ênfase ligeiramente distinta no erro percentual e absoluto.

4.4 MODELAGEM RANDOM FOREST

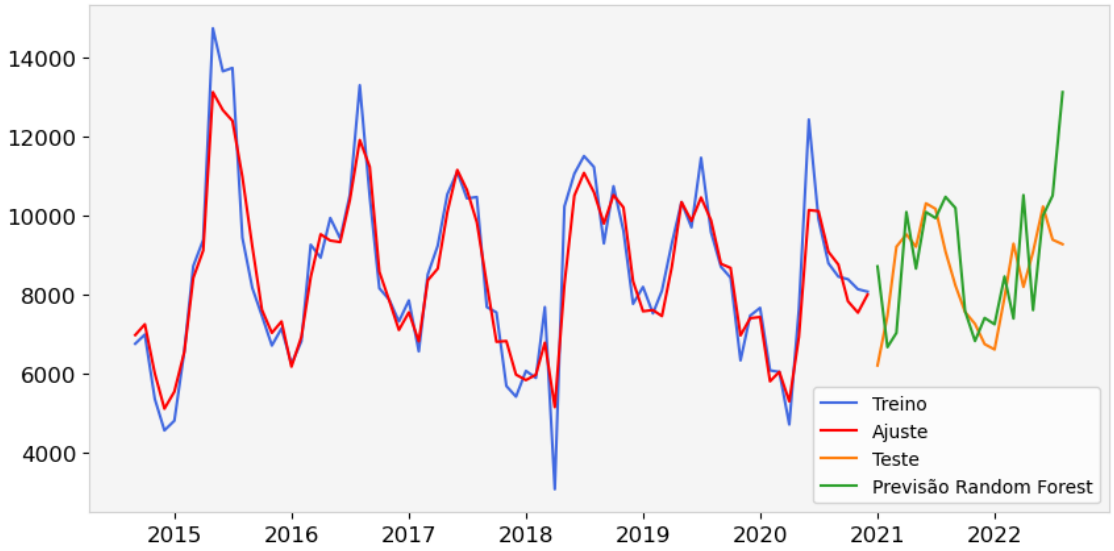
Com os dados ordenados pela coluna “mesAno” e com aplicação da defasagem de período, foi divido a base entre treino e teste, com treino representado pelo ínicio da série, com 80% dos valores, e a teste com 20%, equivalente ao final da série.

Para o treinamento foi instanciado o equivalente a 100 estimadores, que é equivalente a quantidade de arvores de criação criadas para composição da floresta aleatória, este valor foi definido de forma empírica e testado alguns valores superiores, porém não houve melhora significativa na previsão final, por isso decidiu-se em manter.

A partir da Figura 38 é possível verificar visualmente a performance do modelo, onde foi plotado os dados reais de treino e teste, bem como as previsões do modelo. A curva de treino é representada na cor azul, a de teste em laranja, e a previsão em verde. Essa visualização

permite uma análise detalhada sobre a precisão das previsões em comparação com os valores reais.

Figura 38 - Previsão modelo *Random Forest*



Fonte: Elaborado pelo autor (2024).

Observa-se que o modelo apresenta um bom ajuste em diversas regiões ao longo do período analisado. No entanto, algumas discrepâncias são notadas, especialmente nos períodos finais da série temporal. Essas diferenças podem indicar limitações na capacidade do modelo de capturar certas particularidades do conjunto de dados, sugerindo a necessidade de ajustes adicionais ou a consideração de variáveis complementares que possam influenciar os resultados, a fim de melhorar a precisão das previsões.

Em termos de desempenho, o modelo alcançou um MAPE de 14,2%, o que indica um erro percentual médio razoável entre as previsões e os valores reais. O RMSE foi de 1530,4, refletindo a magnitude dos erros em termos absolutos. Esses resultados indicam que, embora o modelo apresente um ajuste satisfatório, há espaço para aprimoramentos, especialmente em períodos em que as previsões não se alinharam tão bem com a realidade.

5 CONCLUSÃO

A previsão de volume de chamadas em *call centers* é um tema fundamental para o dimensionamento de equipes dentro das operações, devido ao alto custo de manutenção de pessoas. Este trabalho explorou diferentes abordagens para a previsão, desde métodos estatísticos clássicos até a aplicação de técnicas de aprendizado de máquina, avaliando seus desempenhos em um estudo de caso real.

Ao trabalhar com modelagens ARIMA, a análise comparativa entre os modelos ARIMA (6,1,6) e (4,0,6) demonstra que, embora ambos apresentem bons resultados, a escolha do modelo ideal depende bastante do objetivo da previsão. O modelo ARIMA(6,1,6) apresentou um desempenho superior em termos de precisão, com valores 2,82 p.p. melhores para o MAPE e 28,6% melhores no RMSE, enquanto o modelo ARIMA (4,0,6) se destacou em termos de ajuste geral aos dados, com o AIC apresentando uma vantagem sutil, mas que contribuiu significativamente para um ajuste no período inicial da série.

A modelagem de aprendizado de máquina, utilizando *Random Forest*, mostrou-se eficiente em capturar padrões complexos da série temporal e obter uma precisão satisfatória, especialmente em cenários com maior variabilidade nos dados. Ao comparar com o melhor modelo ARIMA, o M1, os resultados apresentaram uma performance inferior de 4,7 p.p. para o MAPE e 57,2% no RMSE; em termos relativos, o MAPE foi de 9,6% para a modelagem clássica, enquanto para o *Random Forest* foi de 14,2%. Para o RMSE, foi de 973,5 na modelagem clássica e 1530,4 para o *Random Forest*.

Ao estabelecer um paralelo dos trabalhos estudados presentes na literatura e o desenvolvimento prático, verificou que há proximidade nos resultados das métricas MAPE e RMSE, indicando que é possível obter boas modelagens apenas com uma base contendo informações do período e volumetria.

A aplicação de técnicas avançadas, como o aprendizado de máquina, traz benefícios significativos para a previsão, especialmente em cenários que necessitam ajustes em dados com alta variabilidade. No entanto, o uso de modelos clássicos, como o ARIMA, ainda é importante para a análise da série temporal e a compreensão dos seus padrões. Diferentemente dos trabalhos citados, a modelagem clássica apresentou resultados melhores em comparação com a de aprendizado de máquina considerando a precisão como o principal parâmetro de avaliação.

No que se refere à qualidade do conjunto de dados, foram encontrados alguns períodos com alta variabilidade, tratados como outliers. Esses valores possivelmente limitaram as

modelagens, pois, com um maior range de dados, a qualidade e a possibilidade de captura de padrões aumentam significativamente.

De maneira geral, os resultados obtidos elucidaram que é possível desenvolver uma previsão univariada satisfatória do volume de chamadas em *call centers*, embora, ainda existam possibilidades de melhoria dos modelos, principalmente no que tange à captura de alguns padrões aleatórios, que não foram bem ajustados em ambas as modelagens.

Por fim, identificou-se a oportunidade de realizar estudos incorporando outras variáveis às previsões, como tempo de ligação, taxa de abandono, informações específicas do dia da semana e a análise de cenários com diferentes granularidades temporais, incluindo semanal e diária, além da abordagem mensal já utilizada. Além disso, podem ser incluídas outras abordagens clássicas e de aprendizado de máquina não aplicadas no estudo prático, e modelagens de aprendizado profundo, bem como a realização de testes aplicados a séries temporais em diferentes bases de dados e segmentos.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBRECHT, Tobias; RAUSCH, Teresa Maria. The impact of lead time and model selection on the accuracy of call center arrivals' forecasts. *In*: EUROPEAN CONFERENCE ON INFORMATION SYSTEMS (ECIS), 2020, Marrakech. **Anais** [...]. Marrakech: ECIS. p. 15-17. Disponível em: https://aisel.aisnet.org/ecis2020_rp/19/. Acesso em: 14 maio 2024.

ANTON, Jon. **Call center management by the numbers**. Indiana: Purdue University Press, 1997. Disponível em: https://books.google.com.br/books?id=nAKYXyvTx7IC&printsec=frontcover&hl=pt-BR&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false. Acesso em: 16 ago. 2024.

ATHANASOPOULOS, George; HYNDMAN, Rob J. **Forecasting: principles and practice**. 3. ed. Melbourne: OTexts, 2021. Disponível em: https://otexts.com/fpp3/. Acesso em: 09 jun. 2024.

BALDON, Nicoló. **Time series forecast of call volume in call center using statistical and machine learning methods**. 2019. Dissertation (Master in Computer Science) - KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, Stockholm, 2019.

BERGEVIN, Réal et al. **Call centers for dummies**. 2. ed. Mississauga: John Wiley & Sons, 2010. E-book. Disponível em: https://books.google.com.br/books?id=VJU04_n4W38C. Acesso em: 9 jan. 2024.

BOLFARINE, Heleno. **Elementos de amostragem**. 1. ed. São Paulo: Editora Blucher, 2005. Disponível em: https://app.minhabiblioteca.com.br/reader/books/9788521214991/. Acesso em: 21 ago. 2024.

BOUZADA, Marco Aurélio Carino; SALIBY, Eduardo. Prevendo a demanda de ligações em um call center por meio de um modelo de regressão Múltipla. **Gestão & Produção**, São Carlos, 05 jul. 2009. p. 382-397. Disponível em: https://www.scielo.br/j/gp/a/tYjty3LTnqSCqcR75LLJb5G/?lang=pt. Acesso em: 24 jan. 2024.

BROCKWELL, Peter J.; DAVIS, Richard A. **Introduction to time series and forecasting**. 2. ed. Fort Collins: Springer, 2002. Disponível em: https://otexts.com/fpp3/. Acesso em: 09 jun. 2024.

CASTANHEIRA, Nelson Pereira. **Métodos quantitativos**. 1. ed. Curitiba: Intersaberes, 2013. E-book. Disponível em: https://plataforma.bvirtual.com.br. Acesso em: 27 jun. 2024.
CHACÓN, Henry et al. Forecasting call center arrivals using temporal memory networks and gradient boosting algorithm. *Expert Systems with Applications,* San Antonio, 15 ago. 2023. Disponível em: https://doi.org/10.1016/j.eswa.2023.119983. Acesso em: 1 jun. 2024.

CHANBUNKAEW, Sirithep; THARMMAPHORNPHILAS, Wipawee. Forecasting of incoming calls in a commercial bank service call center. *In*: ICCMS '18: PROCEEDINGS OF THE 10TH INTERNATIONAL CONFERENCE ON COMPUTER MODELING AND SIMULATION, 2018, Sydney. **Anais** [...]. Sydney: Association for Computing Machinery,

2018. p 281-284. Disponível em: https://doi.org/10.1145/3177457.3177498. Acesso em: 1 jun. 2024.

CHAUDHARY, Sana et al. Impact of work load and stress in call center employees: evidence from call center employees. **Pakistan Journal of Humanities and Social Sciences**, Lahore, 1 nov. 2023. p. 160–171. Disponível em: https://internationalrasd.org/journals/index.php/pjhss/article/view/1108. Acesso em: 22 jul. 2024.

CHEN, Tiangi; GUESTRIN, Carlos. Xgboost: a scalable tree boosting system. *In*: KDD '16: PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2016, New York. **Anais** [...]. New York: Association for Computing Machinery, 2016. p. 785-794. Disponível em: https://dl.acm.org/doi/abs/10.1145/2939672.2939785. Acesso em: 10 ago. 2024.

CLEVELAND, Brad; MAYBEN, Julia. **Call center management on fast forward:** succeeding in today's dynamic inbound environment. 1. ed. Maryland: ICMI Inc., 1997. p. 3-98. Disponível em: https://books.google.com.br/books?id=C80Ryymtwz0C&pg=PA1&hl=pt-BR&source=gbs_toc_r&cad=2#v=onepage&q&f=false. Acesso em: 10 ago. 2024.

CLIENTESA. **Empresas brasileiras lideram ranking mundial de atendimento ao cliente**. Disponível em: https://portal.clientesa.com.br/callcenter/empresas-brasileiras-lideram-ranking-mundial-de-atendimento-ao-cliente/. Acesso em: 4 jun. 2024.

DAWSON, Keith. **The call center handbook**: the complete guide to starting, running and improving your call center. 2. ed. New York: Telecom Books and Flatiron Publishing, 1998. cap. 1, p. 1-7.
DEY, Debomit. **ML | Bagging classifier**. Disponível em: https://www.geeksforgeeks.org/ml-bagging-classifier/. Acesso em: 10 jul. 2024.

FACELI, Katti et al. **Inteligência artificial**: uma abordagem de aprendizado de máquina. 2. ed. Rio de Janeiro: LTC, 2021. E-book. Disponível em: https://app.minhabiblioteca.com.br/#/books/9788521637509/. Acesso em: Acesso em: 10 jul. 2024.

FEIGIN, Paul D.; MANDEL, Avishai; NOIMAN, Sivan Aldor. Workload forecasting for a call center: methodology and a case study. **The Annals of Applied Statistics**, Pennsylvania, 1 mar. 2010. v. 3. p. 1403-1447. Disponível em: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-3/issue-4/Workload-forecasting-for-a-call-center--Methodology-and-a/10.1214/09-AOAS255.full. Acesso em: 04 jun. 2024.

FRANZESE, Luis Augusto G. et al. Comparison of call center models. *In*: PROCEEDINGS OF THE 2009 WINTER SIMULATION CONFERENCE (WSC), 2009, Austin. **Anais** [...]. Austin: IEEE, 2009. p. 2963-2970. Disponível em: https://ieeexplore.ieee.org/document/5429228. Acesso em: 30 jul. 2024.

GANS, Noah; KOOLE, Ger; MANDELBAUM, Avishai. Telephone call centers: tutorial, review, and research prospects. **Manufacturing & Service Operations Management**, Pennsylvania, 1 mar. 2003. p. 79-141. Disponível em: https://www.researchgate.net/publication/227448046_Telephone_Call_Centers_Tutorial_Revi

ew_and_Research_Prospects. Acesso em: 08 jul. 2024.

GÉRON, Aurélien. **Mãos à obra: aprendizado de máquina com scikit-learn, keras & tensorflow**: conceitos, ferramentas e técnicas para construção de sistemas inteligentes. 2. ed. Rio de Janeiro: Alta Books, 2021. E-book. Disponível em: https://app.minhabiblioteca.com.br/books/9786555208146. Acesso em: 10 ago. 2024.

GROSSMAN, Thomas A; MEHROTRA, Vijay; SAMUELSON, Douglas A. Call center management. *In*: COCHRAN, James J. et al. **Wiley encyclopedia of operations research and management** science. San Francisco: John Wiley & Sons, 2011. Disponível em: https://www.researchgate.net/publication/319550198_Call_Center_Management/references. Acesso em: 20 jun. 2024.

RIBEIRO, Jairo Moran Carvalho et al. A administração clássica: um estudo aplicado a centrais de atendimento (call center). **Revista de Administração IMED**, Caxias do Sul, v. 5, n. 1, p. 49-58, 2015. Disponível em: https://seer.atitus.edu.br/index.php/raimed/article/view/622/697. Acesso em: 20 jun. 2024.

KAUFMAN, Dora. **Desmistificando a inteligência artificial**.Belo Horizonte: Autêntica, 2022. . Disponível em: https://app.minhabiblioteca.com.br/#/books/9786559281596/. Acesso em: 22 jun. 2024

KOOLE, Ger; SIQIAO, Li; WANG, Qingchen. Predicting call center performance with machine learning. *In*: YANG, Hui; QIU, Robin. **Advances in Service Science:** Proceedings of the 2018 INFORMS International Conference on Service Science. Pennsylvania: Springer, 2019. p. 193-199. Disponível em: https://www.researchgate.net/publication/329979552_Predicting_Call_Center_Performance_with_Machine_Learning_Proceedings_of_the_2018_INFORMS_International_Conference_on_Service_Science. Acesso em: 05 jun. 2024.

LABONE CONSULTORIA E TREINAMENTOS. **Boxplot: veja o que é e qual sua importância**. Disponível em: https://www.laboneconsultoria.com.br/o-que-e-boxplot/. Acesso em: 31 out. 2024.

LIMA, Isaías; PINHEIRO, Carlos A. M.; SANTOS, Flávia A. Oliveira. O. **Inteligência Artificial**. Rio de Janeiro: Elsevier, 2014. p. 1-7. E-book. Disponível em: https://app.minhabiblioteca.com.br/#/books/9788595152724/. Acesso em: 14 maio 2024

MANCINI, Lucas. **Call center: estratégia para vencer**. 2. ed. São Paulo: Summus Editorial, 2006.

MOCELIN, Daniel Gustavo; SILVA, Luís Fernando Santos Corrêa da. O telemarketing e o perfil sócio-ocupacional dos empregados em call centers. **Caderno CRH**, Salvador, v. 21, n. 53, p. 361-383, 2008. Disponível em: https://www.scielo.br/j/ccrh/a/ftZ53zYBPhs39fbBgRr4JfL/?lang=pt. Acesso em: 19 maio 2024.

MOORE, David S.; NOTZ, William I.; FLIGNER, Michael A. **Estatística básica e sua prática**. 9. ed. Rio de Janeiro: LTC, 2023. E-book. Disponível em:

https://app.minhabiblioteca.com.br/#/books/9788521638612/. Acesso em: 23 ago. 2024 e 21 ago. 2024 (referência duplicada no original).

MORETTIN, Pedro A.; BUSSAB, Wilton de O. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017. E-book. Disponível em: https://bookshelf.vitalsource.com/#/books/9788547220228. Acesso em: 10 jul. 2024.

MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de séries temporais**. 3. ed. São Paulo: Editora Blucher, 2018. Disponível em: https://app.minhabiblioteca.com.br/#/books/9788521213529/. Acesso em: 28 jan. 2024.

NORVIG, Peter; RUSSELL, Stuart J. **Inteligência artificial:** Uma Abordagem Moderna. 4. ed. Rio de Janeiro: GEN LTC, 2022. E-book. p. 1-32. Disponível em: https://app.minhabiblioteca.com.br/#/books/9788595159495/. Acesso em: Acesso em: 14 jul. 2024.

PINEDO, Michael; SESHADTI, Sridhar; SHANTHIKUMAR, J. George. Call centers in financial services: strategies, technologies, and operations. *In*: MELNICK, Edward L. et al. **Creating value in financial services**: strategies, operations and technologies. New york: Springer Science+Business Media, 2000. p. 357-388. Disponível em: https://link.springer.com/chapter/10.1007/978-1-4615-4605-4_18#citeas. Acesso em: 22 jul. 2024.

PROBARE. Programa brasileiro de auto-regulamentação do setor de relacionamento (call center / contact center help desk / sac / telemarketing). Disponível em: https://probare.org.br/. Acesso em: 27 maio 2024.

RAMALHO, Luciano. **Python Fluente**. 1. ed. São Paulo: Novatec, 2015. p. 17-18.

RODRÍGUEZ-PÉREZ, Raquel; BAJORATH, Jürgen. Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. **Journal of Computer-Aided Molecular Design**: Incorporating Perspectives in Drug Discovery and Design, 36, 19 mar. 2022. p. 355-362. Disponível em: https://doi.org/10.1007/s10822-022-00442-9. Acesso em: 03 jun. 2024.

ROQUE, Pedro Henrique da Silva et al. Modelagem da previsão do volume de chamadas recebidas por um call center. *In*: SOCIEDADE BRASILEIRA DE AUTOMÁTICA (SBA): XV SIMPÓSIO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE - SBAI, 2021, Rio Grande. **Anais** [...]. Rio Grande: SBAI, 2021. Disponível em: https://doi.org/10.20906/sbai.v1i1.2553. Acesso em: 03 jan. 2024.

SAMARTINI, André; SICSÚ, Abraham Laredo; BARTH, Nelson Lerner. **Técnicas de machine learning**. 1. ed. São Paulo: Editora Blucher, 2023. E-book. Disponível em: https://app.minhabiblioteca.com.br/reader/books/9786555063974/. Acesso em: 09 jul. 2024.

SHARP, Duane. **Call center operation**: design, operation, and maintenance. New Jersey: Elsevier, 2003. Disponível em: https://books.google.com.br/books?id=_P9dZoTEs40C&printsec=frontcover&hl=pt-BR&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false. Acesso em: 24 jul. 2024.

SHUMWAY, Robert H.; STOFFER, David S. **Time series analysis and Its applications**. 4. ed. Pittsburgh: Springer, 2017. Disponível em: http://www.stat.ucla.edu/~frederic/415/S23/tsa4.pdf. Acesso em: 17 ago. 2024.

SIEBEN, Inge et al. Technology, selection, and training in call centers. **ILR Review**, v. 62. New York, 2009. p. 553-572. Disponível em: https://doi.org/10.1177/001979390906200405. Acesso em: 22 jul. 2024.

SOUSA, Alex R. dos Santos et al. **Análise de séries temporais**. 1. ed. Porto Alegre: SAGAH, 2021. E-book. Disponível em: https://app.minhabiblioteca.com.br/#/books/9786556902876/. Acesso em: 27 fev. 2024.

SURASAI, Peerawit. **Time series forecast of call arrivals using machine learning methods**. 2023. Dissertation (Master of Science) - Faculty of Science - Srinakharinwirot University, Bangkok, 2023. Disponível em: http://ir-ithesis.swu.ac.th/dspace/handle/123456789/2577. Acesso em: 27 jun. 2024.

TRIOLA, Mario F. **Introdução à estatística**. 12. ed. Rio de Janeiro: LTC, 2017. E-book. Disponível em: https://app.minhabiblioteca.com.br/reader/books/9788521634256/. Acesso em: 27 fev. 2024.

VASCONCELLOS, Marco Antonio Sandoval de; ALVES, Denisard. Manual de econometria: nível intermediário (2000), São Paulo: Atlas, 2000. Disponível em: https://repositorio.usp.br/item/001053966. Acesso em: 10 set. 2024.

VAUGHN, Robin. **Call centers are helping the economy to rebound**. Disponível em: https://connectionsmagazine.com/article/call-centers-fuel-economic-rebound/. Acesso em: 02 ago. 2024.

ZENONE, Luiz Claudio. **CRM - customer relationship management**: gestão do relacionamento com o cliente e a competitividade empresarial. 1. ed. São Paulo: Novatec Editora, 2007. p. 64-77. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=I0DdXsZ3U0wC&oi=fnd&pg=PA9&dq=CRM&ots=6r2xOzns5H&sig=MzaSYc53CJ17j2d3TnZasiOjQJs#v=onepage&q=CRM&f=false. Acesso em: 30 jul. 2024.