

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA  
UNIDADE DE PÓS-GRADUAÇÃO, EXTENSÃO E PESQUISA  
MESTRADO PROFISSIONAL EM GESTÃO E TECNOLOGIA EM  
SISTEMAS PRODUTIVOS

WILLIAM EMIDIO DE NORONHA

Processo para Diagnóstico e Melhoria de Atendimento em *helpdesk* (PRODIMAH):  
Uma Ferramenta para apoio à Acordo de Nível de Serviços

São Paulo

Mai/2024

WILLIAM EMIDIO DE NORONHA

PROCESSO PARA DIAGNÓSTICO E MELHORIA DE ATENDIMENTO EM *HELPDESK*  
(PRODIMAH): UMA FERRAMENTA PARA APOIO À ACORDO DE NÍVEL DE  
SERVIÇOS

Dissertação apresentada como exigência parcial para obtenção do título de Mestre em Gestão e Tecnologia em Sistemas Produtivos do Centro Estadual de Educação Tecnológica Paula Souza, no Programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos, sob a orientação do Prof. Dr. Marcelo Duduchi Feitosa.

Área de Concentração: Sistemas Produtivos

São Paulo

Maió/2024

Maio/2024

N852p

Noronha, William Emidio de

Processo para diagnóstico e melhoria de atendimento em *helpdesks* (PRODIMAH) : uma ferramenta para apoio à acordo de nível de serviços / William Emidio de Noronha. – São Paulo: CPS, 2024.

120 f. : il.

Orientador: Prof. Dr. Marcelo Duduchi Feitosa

Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos) – Centro Estadual de Educação Tecnológica Paula Souza, 2024.

1. *Helpdesk*. 2. *Big data*. 3. Inteligência artificial. 4. *Machine learning*. 5. Transformação digital I. Feitosa, Marcelo Duduchi. II. Centro Estadual de Educação Tecnológica Paula Souza. III. Título.

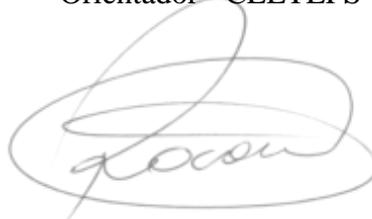
WILLIAM EMIDIO DE NORONHA

Processo para Diagnóstico e Melhoria de Atendimento em *helpdesk* (PRODIMAH):  
Uma Ferramenta para apoio à Acordo de Nível de Serviços



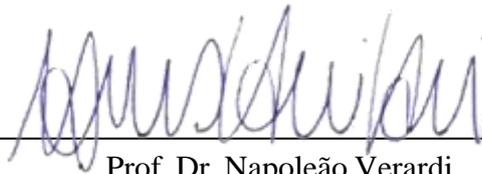
---

Prof. Dr. Marcelo Duduchi Feitosa  
Orientador - CEETEPS



---

Prof. Dr. Ronaldo Celso Messias Correia  
Examinador Externo - UNIVERSIDADE ESTADUAL PAULISTA - UNESP



---

Prof. Dr. Napoleão Verardi  
Galegale  
Examinador Interno - CEETEPS

São Paulo, 14 de maio de 2024

*Dedico este trabalho a todos os  
professores e alunos que participaram  
dessa jornada iniciada em 2022.*

## **AGRADECIMENTOS**

Agradeço primeiramente ao professor Marcelo Duduchi por ter se interessado pelo meu trabalho e por ter me ajudado nos dois anos de curso. Ao professor Napoleão vale não só destacar o agradecimento como também a admiração pela pessoa. Debora Pandolfi e a Sra. Vilma também foram excepcionais profissionais que eu pude conhecer e que fizeram parte dessa intensa jornada de aprendizado. E, certamente, a todos os colegas de classe que se demonstraram sempre solícitos e dispostos a ajudar inclusive, e em especial no final dessa jornada. Agradeço também ao professor Ronaldo que aceitou o convite para participar da banca examinadora e trouxe diversos apontamentos que agregaram valor ao estudo.

Um bom processo é como uma estrada bem pavimentada: ele suaviza a  
jornada e ajuda a alcançar destinos mais distantes.

John C. Maxwell

## RESUMO

Noronha, W. E. **Processo para Diagnóstico e Melhoria de Atendimento em *helpdesk* (PRODIMAH): Uma Ferramenta para apoio à Acordo de Nível de Serviços**. 120 f. Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2024.

A rápida evolução das tecnologias digitais está transformando profundamente as operações de negócios, produtos e processos de gestão nas organizações contemporâneas. Contudo, apesar dessa rápida evolução, a plena realização da transformação digital enfrenta desafios significativos. O sistema de *helpdesk* de Tecnologia da Informação (TI) é vital, especialmente para organizações que dependem fortemente de serviços e recursos de TI, atuando como ponto de contato primário entre a equipe de TI e os usuários. Neste contexto os Acordos de Nível de Serviços desempenham um papel crucial na garantia de um atendimento eficiente ao usuário. Este estudo propõe o PRODIMAH (Processo para Diagnóstico e Melhoria de Atendimento em *helpdesk*), um processo para diagnóstico e melhoria do atendimento do *helpdesk* e apoio na definição de Acordos de Nível de Serviços utilizando técnicas de *machine learning*, contribuindo para o autoconhecimento das organizações e otimização do serviço prestado aos usuários. Para atingir este objetivo, foram definidas as etapas do processo que incluem diagnóstico dos serviços prestados e identificação de viabilidade de aplicação de algoritmos para apoio na definição de tempo de atendimento e melhoria do Acordo de Nível de Serviços (SLA) que foi aplicado para avaliação em uma organização do setor logístico. As contribuições deste estudo incluem uma revisão da literatura sobre o tema e a construção do processo proposto. Na aplicação do processo para avaliação, os resultados iniciais por uma análise de regressão multivariada logística apresentaram taxa de acerto de 95% para chamados atendidos dentro do prazo e 83% para chamados não atendidos dentro do prazo, mas a viabilidade de predição de tempo com os dados existentes não foi identificada. Mesmo assim, o diagnóstico apresentado permitiu aprimorar a prestação de serviços.

**Palavras-chave:** *Helpdesk*, *Big data*, Inteligência artificial, *Machine Learning*, SLA, Transformação Digital

## ABSTRACT

Noronha, W. E. **Process for Diagnosis and Improvement of Helpdesk Service (PRODIMAH): A Tool to support the Service Level Agreement.** 120 f. Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2024.

The rapid evolution of digital technologies is profoundly transforming business operations, products and management processes in contemporary organizations. However, fully realizing digital transformation faces significant challenges. The Information Technology helpdesk system is vital, especially for institutions that rely heavily on Information Technology (IT) services and resources, acting as the primary point of contact between the IT team and users. In this context, Service Level Agreements play a crucial role in ensuring efficient service to the user. This study proposes PRODIMAH (Process for Diagnosis and Improvement of Service in Helpdesks), a process for diagnosing and improving helpdesk service and assertiveness in defining Service Level Agreements (SLA) using machine learning techniques, contributing to the self-knowledge of organizations and optimization of the service provided. to users. The stages of the process were defined, including diagnosis of the services provided and identification of the feasibility of applying algorithms for greater assertiveness in defining service time and improving the Service Level Agreement (SLA), applied for evaluation in an organization in the logistics sector. The contributions of this study include a review of the literature on the topic and the construction of the tool. When applying the process for evaluation, the initial results from a multivariate logistic regression analysis showed a hit rate of 95% for calls answered within the deadline and 83% for calls not answered within the deadline, but the feasibility of predicting time with the existing data was not identified. Even so, the diagnosis presented improved the provision of services based.

**Keywords:** Helpdesk, Big Data, Artificial Intelligence, Machine Learning, SLA, Digital Transformation.

## LISTA DE QUADROS

Quadro 1 - Benefícios do uso de TD .....	28
Quadro 2 - Definições de <i>Big Data</i> .....	31
Quadro 3 - Representação de cada V.....	32
Quadro 4 - Strings de busca.....	52
Quadro 5 - Web of Science - Palavras de interesse .....	52
Quadro 6 - Scopus - Palavras de interesse.....	53
Quadro 7 -Análise dos artigos selecionados da revisão sistemática.....	56
Quadro 8 - Técnicas utilizadas e suas quantidades.....	58
Quadro 9 - Variação no tempo decorrido para atendimento .....	84

## LISTA DE GRÁFICOS

Gráfico 1 - Coluna Departamento com percentual .....	78
Gráfico 2 - Coluna Categoria com percentual .....	79
Gráfico 3 - Subcategoria por percentual .....	79
Gráfico 4 - Criado por com mais de 80 solicitações em percentual .....	80
Gráfico 5 - Criado por até 40 solicitações .....	80
Gráfico 6 - Técnico com mais de 50 solicitações .....	81
Gráfico 7 - Local em percentual .....	81
Gráfico 8 - Primeira resposta atrasada em percentual .....	82
Gráfico 9 - Categoria com mais de 50 horas .....	82
Gráfico 10 - Criado por com mais de 50 horas .....	83
Gráfico 11 - Técnico com tempo de atendimento superior a 50 horas.....	83
Gráfico 12 - Seis primeiros da subcategoria com mais de 50 horas.....	84
Gráfico 13 - Coluna dias da semana.....	85

## LISTA DE FIGURAS

Figura 1 – Hierarquia das Funções de um SLA.....	23
Figura 2 – Proposta de um Modelo de Acordo de Nível de Serviço.....	25
Figura 3 - Os 7 V's do <i>Big Data</i> .....	32
Figura 4 - Modelo de construção e teste de ML.....	38
Figura 5 - Etapas propostas para a condução de pesquisas utilizando DSR.....	47
Figura 6 - Macro etapas da revisão sistemática da literatura.....	51
Figura 7 - Fluxo para revisões sistemáticas atualizadas que incluíram apenas buscas em bases de dados e registros .....	54
Figura 8 – Técnica utilizada versus quantidade.....	59
Figura 9 - Proposta para resolução .....	63
Figura 10 - Versão utilizada do Python .....	66
Figura 11 - Bibliotecas utilizadas .....	67
Figura 12 - Colunas do <i>dataset</i> .....	67
Figura 13 - Análise inicial da coluna Departamento .....	70
Figura 14 - Tempo médio por Departamento .....	70
Figura 15 - Análise inicial da coluna Categoria .....	71
Figura 16 - Tempo médio por Departamento .....	71
Figura 17 – Análise inicial Subcategoria.....	72
Figura 18 - Tempo médio por Subcategoria.....	72
Figura 19 - Análise inicial Criado por .....	73
Figura 20 - Seis maiores tempo médio Criado por.....	74
Figura 21 - Análise inicial Técnico .....	74
Figura 22 - Seis maiores médias de tempo por técnico .....	74
Figura 23 - Análise inicial da coluna Local.....	75
Figura 24 - Seis maiores médias de tempo por local.....	75
Figura 25 - Análise inicial da coluna primeira resposta atrasada .....	76
Figura 26 - Análise inicial da coluna Status do vencimento .....	76
Figura 28 - Exemplo de uso do <i>Label encoder</i> .....	86
Figura 29 - <i>Dataset</i> para pré-processamento .....	87
Figura 30 - Atribuição das <i>features</i> a variável X.....	87
Figura 31 - Variável alvo.....	87
Figura 32 - Divisão treino / teste .....	87
Figura 33 - Exibição dos conjuntos resultantes .....	88
Figura 34 - Resultado regressão logística.....	88
Figura 35 - Resultado matriz de confusão .....	88
Figura 36 - Modelo de regressão logística .....	89
Figura 37 - Variação das variáveis da lista features .....	89

## LISTA DE SIGLAS

BM's	.....	Business Models
EDA	.....	Exploratory Data Analysis
GSTI	.....	Gerenciamento de Serviços de Tecnologia da Informação
IA	.....	Inteligência artificial
ML	.....	Machine Learning
NLP	.....	Natural Language Processing
QoS	.....	Qualidade do Serviço
SLA	.....	Service Level Agreement
SVM	.....	Support Vector Machine
TD	.....	Transformação Digital
TI	.....	Tecnologia da Informação

## SUMÁRIO

<b>INTRODUÇÃO .....</b>	<b>15</b>
<b>1. FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>20</b>
1.1 Gestão dos serviços de TI .....	20
1.1.1 Helpdesk .....	22
1.2 Acordo de Nível de Serviços.....	23
1.3 Transformação digital.....	27
1.4 Big Data .....	30
1.5 Ciência de dados .....	33
1.6 Inteligência artificial.....	34
1.7 Machine learning.....	36
1.7.1 Limpeza dos dados .....	39
1.7.2 Análise exploratória .....	40
1.7.3 Engenharia de atributos .....	41
1.7.4 Pré-processamento .....	42
<b>2. METODOLOGIA.....</b>	<b>45</b>
<b>3. REVISÃO SISTEMÁTICA DA LITERATURA .....</b>	<b>51</b>
<b>4. PROCESSO PROPOSTO (PRODIMAH) .....</b>	<b>61</b>
4.1 Identificação dos estudos e configuração das classes de problemas .....	61
4.2 Proposição do artefato para resolver o problema .....	61
4.3 Projeto e desenvolvimento do artefato selecionado.....	62
<b>5. RESULTADOS .....</b>	<b>65</b>
5.1 Aplicação do processo proposto .....	65
5.1.1 Análise inicial dos dados.....	69
5.1.2 Processo de limpeza dos dados .....	78
5.1.3 Análise exploratória dos dados .....	78
5.1.4 Engenharia de atributos .....	85
5.1.5 Pré-processamento .....	86
5.1.6 Avaliação do processo por profissionais especialistas .....	90
5.1.7 Explicitação das aprendizagens .....	93
<b>CONCLUSÃO.....</b>	<b>97</b>
<b>REFERÊNCIAS .....</b>	<b>99</b>
<b>APÊNDICE .....</b>	<b>104</b>

## INTRODUÇÃO

Em todo o mundo, indústrias têm conduzido uma série de iniciativas para explorar novas tecnologias digitais e seus benefícios que frequentemente envolvem transformações nas principais operações de negócios, afeta produtos, processos e conceitos de gestão (MATT; HESS; BENLIAN, 2015).

Essa série de iniciativas é conhecida como transformação digital. Destaca-se a grande velocidade com que ela aconteceu nessas últimas duas décadas enfatizando a importância e o poder que os dados hoje representam (GUNAY, 2018).

Todas essas mudanças tornam necessário o gerenciamento de serviços de tecnologia da informação (GSTI) que é considerado uma coleção de frameworks que dão suporte às organizações que gerenciam serviços de tecnologia. Ainda que a implementação desses tipos de frameworks tem aumentado constantemente no domínio do provedor de serviços de Tecnologia da informação (TI), os principais conceitos de GSTI ainda são novos para muitas organizações que ainda têm dificuldade em vários contextos neste domínio, principalmente durante a implementação (SERRANO *et al.*, 2021).

Macleane e Titah (2023) apontam investimentos significativos em TI feitos pelas organizações, incluindo programas de transformação digital, com o objetivo de aprimorar a qualidade e a entrega de serviços para criar maior valor para as partes interessadas. No entanto, para realizar investimentos em TI visando a efetiva transformação digital é necessário, entre outros fatores, um domínio sólido da função de TI.

O GSTI emergiu como uma abordagem consistente, amplamente implementada para gerenciar a função de TI de uma organização, cujo objetivo é o de obter benefícios operacionais e estratégicos (MACLEAN; TITAH, 2023).

Considerando GSTI nesse contexto Al-Hawari e Barham (2021) trazem que um sistema de *helpdesk* de TI tornou-se necessidade vital para grandes instituições que dependem fortemente de serviços e recursos de TI, visto que serve como ponto único de contato entre a equipe de TI e usuários sobre os serviços solicitados e relatados com algum problema. Além disso, facilita o gerenciamento de *tickets* de TI e permite a automação de tarefas diárias de TI como, por exemplo, atribuição de *tickets* para agentes de serviço e notificações por e-mail para

partes relacionadas. Pode também ajudar a definir e minimizar as atividades que compreendem os diferentes processos de negócios.

Steinbacher *et al.*, (2023), nesse contexto, destacam a importância de classificar *tickets* de suporte ao cliente de acordo com os critérios desejados no GSTI. A classificação precisa permite que os agentes de suporte reutilizem soluções, o que, por sua vez, reduz o tempo de resolução de *tickets* e aumenta a satisfação do cliente.

Al-Hawari e Barham (2021) acrescentam e apontam que a GSTI possibilita a avaliação do desempenho geral do departamento de TI com base em relatórios gerados e indicadores-chave de desempenho avaliados. Além disso, o uso de software que esses serviços dão suporte de uso em uma organização leva ao aumento da produtividade, melhor qualidade de serviço e melhor satisfação do cliente.

Nessa dinâmica de transformação temos de acordo com Ahmed *et al.*, 2023 que. o amadurecimento da análise de *Big Data* viabilizando às empresas aprimorar seus processos de negócio usando uma abordagem orientada por dados. Desde varejistas, até grandes *players* do comércio eletrônico, passando por empresas de bioinformática, informática de saúde entre outros estão pensando em extrair dados de negócios em informações para tomar melhores decisões a fim de gerar valor para os negócios.

Continuando o raciocínio, temos, a falta de engenheiros e analistas de dados torna mais difícil para as empresas aplicarem a análise de *Big Data* em seus dados de negócios com o objetivo de aprimorar seus processos operacionais. Em última análise, esses processos podem ser automatizados para reduzir erros humanos (AHMED *et al.*, 2023).

Lu, Cairns e Smith (2020) apontam que uma grande quantidade de dados complexos está sendo gerada no ambiente de negócios, o que permite suporte para a tomada de decisão por meio do processamento de informações e geração de *insights*.

Nesse contexto Cybulski e Scheepers (2021) apontam que o campo da ciência de dados surgiu nos últimos anos, com base nos avanços em estatística computacional, *Machine Learning*, Inteligência Artificial (IA) e *Big Data*. Complementam afirmando que as organizações modernas estão imersas em dados e estão se voltando para a ciência de dados para resolver uma variedade de problemas de negócios.

Considerando a inteligência artificial (IA) como consequência da TD, Bhattacharya (2019) afirma que a IA caminhou de sua torre de marfim acadêmica em direção ao mundo comercial, e as consequências despertaram interesse na academia e na indústria.

Gunay *et al.*, (2018) em seu trabalho concluem afirmando que em tempos atuais, o *Big Data* e as ferramentas de *machine learning* tornaram-se muito populares tanto na literatura quanto na indústria. As pessoas usam *Machine Learning* para obter informações significativas do *Big Data* o que traz resultados de planejamento valiosos.

Rossi, Rubattino e Viscusi, (2019) acrescentam e reforçam que *Big Data* e a análise de dados receberam grande atenção de profissionais e acadêmicos, hoje em dia, representando um recurso chave para o interesse renovado em IA, especialmente para técnicas de ML.

Complementando o raciocínio Miklosik *et al.*, (2019) trazem que a expansão tecnológica exponencial cria oportunidades de vantagem competitiva ao aplicar novas abordagens orientadas a dados. *Machine learning* pode prever futuros desenvolvimentos e apoiar a tomada de decisões extraindo *insights* de grandes quantidades de dados gerados. Essa funcionalidade impacta e agiliza muito o processo de tomada de decisão estratégica das organizações.

Tangenciando o uso do *Big Data* Lee e Chien (2022) colaboram e apontam que impulsionada pela migração contínua para a indústria 4.0, a crescente adoção de IA, uso de *Big Data*, computação em nuvem, internet das coisas e robótica capacitaram a fabricação inteligente e a transformação digital. No entanto, o aumento da aplicação de técnicas de ML e ciência de dados apresenta uma série de questões processuais, incluindo aquelas que envolvem dados, suposições, metodologias e condições aplicáveis. O que na prática aumenta as dificuldades para implementação, especialmente associada às características de fabricação e conhecimento do domínio.

A partir do cenário apresentado, verifica-se uma dificuldade que várias empresas apresentam em acompanhar as inovações tecnológicas. Muitas empresas se inserem em novas iniciativas de forma prematura, investem em novas tecnologias que podem não ser ideais para o seu modelo de negócio. Este ponto nos leva concluir que seria necessário realizar abordagens baseadas em pesquisas mais criteriosas.

Diante desse momento transformativo, em especial a partir da consciência da importância que os dados assumiram também e, em especial, nas centrais de suporte, surgiu a ideia do presente trabalho de construir um processo para diagnóstico e melhoria dos serviços de *helpdesk* utilizando recursos da transformação digital e identificando a viabilidade de uso de técnicas de ML para auxiliar organizações a melhor definirem o SLA (*Service Level*

*Agreement*.) a partir de predição de tempo de atendimento de *helpdesk*, tirando melhor proveito do uso dessas novas tecnologias.

Um dos fatores motivadores para a realização deste trabalho é a experiência do autor na atuação como coordenador de equipes de *helpdesk*, onde observou a dificuldade recorrente em atender ao Acordo de Nível de Serviço (SLA) em diversas empresas. Também são elementos motivadores deste estudo, a crescente demanda das organizações por uma melhor adequação de recursos nas centrais de suporte ao usuário e a oportunidade de utilizar técnicas avançadas de *Machine Learning* (ML) para a predições em atendimentos. Entende-se esta como uma pesquisa necessária, uma vez que, de acordo com Dresch *et al.*, (2015) pesquisa necessária é aquela que conjuga o rigor teórico-metodológico e a utilidade prática para a sociedade.

A pergunta que se pretende responder na presente pesquisa é “Quais as etapas necessárias para um processo que funcione como uma ferramenta de apoio com o uso de ML que possibilite um melhor atendimento de *helpdesk* e melhoria da assertividade do SLA?”

O objetivo geral deste trabalho é construir um processo de apoio que inclui o uso de estatística descritiva e ML para compreender e estruturar melhor o atendimento de *helpdesk* e a assertividade ao SLA.

São objetivos específicos dessa pesquisa:

- Revisão da literatura sobre acordos de níveis de serviços, *helpdesk* e uso de ML, para identificar a melhor forma de utilizar esses recursos em favor de melhorias em SLA;
- Identificação de possíveis processos existentes;
- Construção do processo proposto;
- Avaliação do processo proposto.

Como contribuição deste trabalho temos a revisão bibliográfica sobre o assunto e os estudos para predição relacionadas ao atendimento de chamados que oferecem uma abordagem inovadora para melhorar a eficiência dos serviços de *helpdesk*, e o processo proposto denominado PRODIMAH. O uso do PRODIMAH possibilita um melhor planejamento do SLA, com identificação precisa de objetivos, requisitos e designação clara dos responsáveis. A factibilidade do processo é destacada pela importância da análise da base de dados para identificar tendências e propor melhorias ao SLA. Foram evidenciados os desvios existentes e foram sugeridas melhorias. A usabilidade do processo enfatiza a divulgação e geração de

expectativas sobre prazos dos serviços, além da necessidade de alinhamento de expectativas entre as partes envolvidas no atendimento.

O estudo está ligado a linha de pesquisa Sistemas de Informação e Tecnologias Digitais e ao projeto Tecnologias Digitais em Sistemas Produtivos, do programa de mestrado profissional do Centro Paula Souza. Relacionado especificamente ao tema Tecnologias Digitais na Indústria e Serviços, onde o estudo do tema transformação digital se conecta ao momento de posse, empoderamento e uso de novas tecnologias pelas organizações.

Este trabalho está organizado em capítulos, começando com a exploração dos conceitos no referencial teórico no Capítulo 1, a metodologia adotada é detalhada no capítulo 2. O capítulo 3 apresenta a revisão sistemática da literatura e seus respectivos resultados. No capítulo 4 são discutidos os resultados obtidos, acompanhados das aprendizagens extraídas. Finalmente, a conclusão do trabalho é apresentada no último capítulo.

## 1. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são explorados os conceitos de Gerenciamento de Serviços de Tecnologia da Informação (GSTI), Transformação Digital (TD), *Big Data*, Ciência de Dados, Inteligência Artificial (IA) e *Machine Learning* (ML), destacando suas interconexões e importância na construção das etapas do processo para servir como ferramenta de apoio ao atendimento de SLA.

Tangendo à TD ela será tratada devido a sua integração e ao aprimoramento dos processos digitais em uma organização, fornecendo a base necessária para adotar tecnologias avançadas. O estudo de *Big Data* é necessário pois envolve o gerenciamento e a análise de grandes conjuntos de dados possibilitando a análise de dados históricos de atendimento para estimar o tempo de atendimento possibilitando obter *insights* e tomar decisões mais assertivas. A Ciência de Dados, viabiliza a aplicação técnicas estatísticas, uso de recursos computacionais tal como o uso de algoritmos de ML para desenvolver modelos de predição de tempo de atendimento.

Focada no desenvolvimento de sistemas inteligentes capazes de realizar tarefas complexas, como a predição de tempo de atendimento a IA aprende com os dados históricos e adapta suas previsões com base em mudanças nos padrões de atendimento.

Uma subárea da Inteligência Artificial, o ML, desenvolve algoritmos e modelos que permitem que os sistemas aprendam e melhorem com os dados, permitindo previsões cada vez mais precisas sobre o tempo de atendimento.

### 1.1 Gestão dos serviços de TI

De acordo com Silva (2014), estamos vivendo constantes mudanças e a sociedade está cada vez mais esclarecida dos seus direitos e deveres como cidadão, com informações mais acessíveis e instantâneas. Esse fator torna os consumidores mais exigentes no momento da escolha de um serviço. Com isso, a empresa deve transmitir credibilidade e qualidade naquilo que é oferecido. Caso contrário, não conseguirá manter-se no mercado por muito tempo. Por isso, é necessário à busca pela inovação e a melhoria contínua dos seus processos.

O setor de serviços vem indicando crescimentos expressivos nos últimos anos e tendo uma grande contribuição para o crescimento da economia, mas as empresas precisam se preparar para este cenário, com serviços de qualidade e condições de preços acessíveis aos consumidores (SILVA, 2014).

De acordo com Al-Asmoery *et al.*, (2021) a GSTI é um processo de gerenciamento de serviço que adota uma abordagem específica para o serviço de TI. O GSTI nos fornece uma maneira de implementar e gerenciar serviços de TI de qualidade por meio de pessoas, processos, e tecnologia de suporte que atende às necessidades de negócios.

TI é uma combinação da própria tecnologia e uma coleção de serviços que garantem a implementação efetiva da TI geral em uma organização. Como outros serviços, o GSTI tornou-se um assunto global porque o gerenciamento de TI só é possível por meio de protocolos eficientes e eficazes. Um sistema de entrega de GSTI bem implementado melhora a qualidade dos serviços de TI, o que eventualmente aumenta a capacidade e a produção geral da organização (SARWAR *et al.*, 2023).

O de acordo com Mao, Zhang e Tang (2021) o GSTI, impulsionado pela inteligência artificial (IA), mudará substancialmente os modelos de negócios e as operações corporativas, iniciando uma onda de transformação da IA corporativa.

O empoderamento da IA trouxe nova vitalidade e desafios para as capacidades de GSTI, e as empresas devem começar a pensar em como a capacidade de GSTI com IA pode ser reativada para moldar a resiliência dos modelos de negócios e atender às necessidades de sobrevivência em ambientes complexos e em constante mudança (MAO; ZHANG; TANG, 2021).

Zuev *et al.*, (2018) apontam que o GSTI é uma variedade de atividades voltadas para a manutenção da infraestrutura de TI. Portanto, é considerado uma atividade importante para qualquer organização. Sendo o tempo de resolução de incidentes um indicador chave para GSTI. A investigação acerca dos diversos fatores que influenciam o tempo de resolução de incidentes e a eventual aplicação de modelos de ML em GSTI para isso permite melhorias significativas na experiência do cliente e lidando com problemas de forma mais eficiente, diminuindo os esforços dos agentes de *helpdesk* e reduzindo os custos de serviço.

O contexto contemporâneo caracterizado por constantes mudanças e uma sociedade mais informada tornou os consumidores mais exigentes também na escolha de serviços. Temos que a busca pela inovação e melhoria contínua dos processos torna-se essencial para a

permanência das organizações no mercado. O setor de serviços, especialmente o de TI, experimenta um crescimento significativo, onde podemos destacar a importância do GSTI.

O GSTI, portanto, surge como uma atividade importante não só para a manutenção eficiente da infraestrutura de TI e sucesso organizacional em ambientes complexos e dinâmicos, mas também para o *helpdesk*, uma função fundamental dentro do processo de gestão dos serviços.

### 1.1.1 *Helpdesk*

De acordo com Shanmugalingam *et al.*, (2019) *apud* Park *et al.*, (2019) em uma organização, no suporte de TI a operação de *helpdesk* é uma unidade importante que lida com os serviços de TI de uma empresa. Muitas organizações de grande porte tem uma equipe de suporte de TI abrangente para lidar com o engajamento e solicitações com os funcionários 24x7. Como qualquer tarefa rotineira, a maioria dos processos do suporte são considerados de natureza repetitiva sendo que alguns ocorrem diariamente e outros ocorrem com menos frequência e, portanto, muitos engenheiros de suporte e agentes gastam tempo com essas tarefa repetitiva, como as de inserir informações em um aplicativo.

O *software helpdesk* ou sistema *helpdesk* fornece um ponto de suporte e contato entre o departamento de TI provedor e os usuários finais. Ter esta ferramenta permite tomar decisões diárias que influenciam o forma como os incidentes tecnológicos são resolvidos e, assim, prevenir que os processos dos diferentes áreas sejam afetadas por um longo período (GALLARDO *et al.*, 2018).

Em seu trabalho Al-Hawari e Barham (2021), trazem que um sistema de *helpdesk* atua como um único ponto de contato entre os usuários e a equipe de TI. E complementam apontando que em sua pesquisa foi utilizado um modelo de ML de classificação de tíquetes para associar um tíquete de suporte técnico com seu serviço correto desde o início e, portanto, minimizar o tempo de resolução de tickets, economizar recursos humanos, e aumentar a satisfação do usuário.

Em seu estudo Andrews *et al.*, (2016) apontam e trazem, um modelo de custo para operações de *helpdesk* os incidentes previstos com os custos de mão de obra.

Sistemas de *helpdesk* modernos podem incorporar modelos de ML para, por exemplo, classificar tickets, minimizando o tempo de resolução, otimizar o uso de recursos humanos e aumentar a satisfação do usuário. Por meio de uma relação sinérgica com o *helpdesk* o Acordo

de Nível de Serviços (SLA), próximo tópico discutido neste trabalho, fornece o quadro estrutural para garantir que o *helpdesk* atenda aos padrões de desempenho desejados e forneça serviços de suporte de alta qualidade aos usuários finais sendo que a integração eficaz desses dois elementos é essencial para a eficiência operacional e a satisfação do cliente na gestão de serviços de TI. Considerando a importância dessa sinergia a discussão é direcionada para o próximo tópico.

## 1.2 Acordo de Nível de Serviços

Um SLA é um tipo especial de contrato legal que vincula um fornecedor a seus clientes onde o fornecedor se compromete a fornecer determinados serviços em troca para determinado pagamento dos clientes (UPADHYAY *et al.*, 2021).

Um SLA atua como um contrato juridicamente vinculativo que obriga os provedores de serviços a cumprirem com a qualidade de serviço (QoS) prometida sendo um método contratual que especifica e rege a prestação de serviços entre provedores de serviços, consumidores e outras partes (ALZUBAIDI *et al.*, 2019).

De acordo com Forconi (2016) apud Lee e Bem-Natan (2002), conseguir implementar um SLA que funcione bem para cliente e provedor é uma tarefa bastante difícil porque ambas as partes costumam ter uma percepção diferente das funções de um SLA. As funções mais frequentemente encontradas em um contrato deste tipo são seis e seguem uma hierarquia conforme mostra a figura 1.

Figura 1 – Hierarquia das Funções de um SLA



Fonte: Forconi (2016) apud Lee e Ben-Natan (2002)

Prosseguindo com o raciocínio Forconi (2016) *apud* Lee e Bem-Natan (2002), trazem que um SLA bem desenvolvido deve ter funções e responsabilidades definidas da maneira mais clara e concreta possível, considerando responsabilidades e recursos disponíveis para ambas as partes. Sua elaboração deve conter informações suficientes para determinar a expectativa do cliente quanto à entrega do produto ou serviço por parte do fornecedor. Sendo um documento que possibilita administrar a execução do serviço, assegurando a entrega periódica e o desempenho contínuo do serviço. Formalizadas as exigências para os níveis de serviços elas devem ser atingidas, e é de extrema importância para o provedor e para o cliente, que estes níveis de serviços sejam medidos. Por meio da comunicação entre fornecedor e cliente expressam suas necessidades e suas expectativas de desempenho relativo ao progresso da atividade a fim de que haja melhora do SLA e consequente desempenho do serviço entregue. No topo da pirâmide o ROI que significa Retorno sobre o Investimento busca demonstrar o grau de eficácia, podendo ser evidenciado o desempenho tanto da qualidade dos produtos quanto dos serviços prestados.

Violações de SLA acontecem no mundo real. Uma violação do SLA representa a falha na garantia de um serviço, o que leva a consequências indesejadas, como pagamento de multas, redução da margem de lucro, degradação da reputação, rotatividade de clientes e interrupções do serviço (ZENG *et al.*, 2021).

A existência de opções restritas de SLA, que normalmente correspondem a alguns níveis de serviço, pode fazer com que um cliente aceite um serviço que pode não responder de forma eficaz às suas necessidades. Do ponto de vista do provedor de serviços, também é um modelo de negócios abaixo do ideal, com a capacidade sendo reservada para clientes que não a usarão e, posteriormente, indisponível para clientes que o usariam. Portanto, defende-se o uso de SLA personalizado para evitar tais situações, que podem ser configuradas idealmente sem a ajuda de um operador humano (PEOPLES *et al.*, 2023).

Khan *et al.*, (2022) trazem que o SLA gerencia o relacionamento entre provedores de serviços e consumidores e complementam apontando que o SLA é uma parte integrante e crítica dos fornecedores de TI e contratos de comunicação da era moderna.

Com a proposta de disponibilizar um modelo de SLA Gomes *et al.*, (2021) trazem uma proposta cujo objetivo é servir de guia para que organizações definam SLA, e este modelo, além de se correlacionar com as ISOs 9000 e 2000 pode ser dividido em três grandes fases.

Para eles o planejamento do acordo de nível de serviço, primeiramente deve identificar os objetivos da empresa que a levaram à implantação da gestão de determinados serviços de TI por Acordo de Nível de Serviço, um diagnóstico da situação atual e a definição de planos relevantes. Apresentam também uma proposta de um Modelo de Acordo de Nível de Serviço conforme Figura 2.

Figura 2 – Proposta de um Modelo de Acordo de Nível de Serviço

Fase	Etapa	Ítem
Planejamento do SLA	Identificação de objetivos e requisitos	
	Identificação dos gestores	
	Diagnóstico da situação atual	Infra-estrutura dos serviços
		Serviços Nível atual dos serviços
		Utilização atual de recursos
		Hardware e Software
	Planos relevantes para o SLA	
Definição do SLA	Detalhes sobre o SLA	Objetivo do acordo
		Modalidade do acordo
		Datas de revisão
		Responsabilidades
		Coleção da definição dos termos usados no acordo
	Detalhes dos serviços	Identificação
		Definição e escopo
		Pré-requisitos para a execução
		Níveis de serviço contratados
		Volume
	Remuneração	
Definição da gestão do SLA	Métricas	
	Acompanhamento dos serviços	
	Divulgação dos resultados	

Fonte: Adaptado de Gomes *et al.*, (2021)

É necessário que sejam levantados os requisitos que vão permitir à empresa alcançar os objetivos identificados, respeitando-se suas políticas e estratégias, seus processos e produtos, assim como seus limites financeiros.

Deve-se também conhecer o mercado, visando identificar o que está sendo praticado no mercado com relação à contratação dos serviços requeridos e padrões adotados, para que seja possível fazer uma avaliação considerando os provedores de serviços atuais e potenciais candidatos.

Uma vez definidos os responsáveis, deve-se providenciar um diagnóstico da situação atual, que deve incluir a infraestrutura necessária ao serviço para quantificar o que será contratado considerando a infraestrutura, tipos de serviço e situação atual para que se possa avaliar o nível atual de serviços, a utilização de recursos e o *hardware/software* utilizados. Esta avaliação deve ser feita a partir dos dados existentes e deve incluir: uma lista de serviços, com as métricas e indicadores considerados, os custos envolvidos para manter o nível de serviço atual.

Gomes *et al.*, (2021) trazem também que riscos devem ter um plano que defina regras que serão válidas e procedimentos que serão adotados no caso de sua ocorrência. Os mais relevantes no contexto do SLA são:

- Plano de Transição – define regras a serem aplicadas durante o período pré-determinado de passagem do serviço para a responsabilidade do fornecedor;
- Plano de Qualidade – define requisitos para notificação, guias para referência e resposta, padrão de documentação e padrão para a execução do serviço;
- Plano de Segurança – define as regras a serem seguidas para cada tipo de serviço, explicitando penalidades caso ocorra algo fora do esperado;
- Plano de Gestão – estabelece a severidade dos problemas, como se dará a sua resolução, qual o processo de gestão de chamadas, horários de cobertura do *helpdesk*, como se dará o rastreamento de chamadas etc.

Na fase de definição do SLA, de acordo com Gomes *et al.*, (2021), deve-se definir claramente o escopo, os níveis de serviço contratados e os valores envolvidos. O cliente deve definir os detalhes sobre o acordo de nível de serviço considerando seu objetivo, modalidade

(interna, externa ou mista), datas das revisões, responsabilidades e deve ser definido um glossário dos termos usados no acordo.

Os detalhes do serviço incluem o nome do serviço, escopo, descrição do serviço, um indicativo apontando se o cliente pode contratar esse serviço com outro provedor ou não, definições do que está e não está incluso no serviço, pré-requisitos para a execução devem especificar o que deve estar disponível, com que período de antecedência e regras de uso do serviço, dentre outros. Apontamentos que tange às necessidades, desejos, prioridades, indicativos sobre o nível do serviço, comunicação sobre desvios, devem ser claros, assim como a conduta na resolução de problemas. O volume do serviço a ser provido, define parâmetros que possibilitem quantificar o que foi contratado. Deve indicar como fazer os ajustes no caso discrepâncias em relação ao que foi contratado. Com relação à remuneração, deve determinar os custos de prover os níveis de serviço desejados, as taxas a serem aplicadas, penalidades, multas e bonificação.

Concluindo sua proposta de um modelo de acordo de nível de serviço Gomes *et al.*, (2021) trazem que na fase da definição da gestão do SLA, o desempenho na execução dos serviços contratados sob o SLA precisa ser mensurado para que este possa ser acompanhado, seus resultados divulgados e possibilitar propostas de evolução.

Guedes *et al.*, (2021) trazem que a proposição de um modelo de SLA para controle do processo de manutenção operacional da gestão de SLA devem ter como características centrais a confiabilidade, portabilidade e disponibilidade de valores pré-definidos requeridos, a avaliação contínua de desempenho tal como análises de tendências dos critérios para mensurar sua utilidade.

Tendo abordado os conceitos principais de SLA o presente trabalho passa a discorrer sobre os aspectos específicos que podem contribuir para a transformação digital na GSTI.

### **1.3 Transformação digital**

Considerando importância da contextualização do tema, esse tópico do referencial teórico traz o conceito de Transformação Digital (TD) correlacionando-o ao momento de mudanças vivenciado pelas organizações.

As estratégias de TD se concentram na transformação de produtos, processos e organização devido às novas tecnologias. Seu alcance é mais amplamente projetado e explicitamente inclui atividades digitais (MATT; HESS; BENLIAN, 2015).

De acordo com Mugge *et al.*, (2020) a TD está remodelando segmentos e setores inteiros: comunicações, varejo e, cada vez mais, saúde, medicina, agricultura e manufatura, sendo a TD uma das principais preocupações dos líderes seniores em todo o mundo.

Eickemeyer *et al.*, (2021) apresentam os vários impactos positivos que a crescente implementação de tecnologias digitais tem beneficiado as organizações. No entanto, algumas organizações, muitas vezes, buscam implementar tendências tecnológicas sem preparação suficiente e prestam atenção insuficiente aos fatores humanos envolvidos na digitalização.

A TD proporciona uma grande mudança no desempenho de uma organização centrada em tecnologias transformadoras como *Internet of Things* (IoT), computação em nuvem, análise de dados, inteligência artificial e *blockchain*, que estão entre os tipos mais importantes de tecnologias transformadoras (CAVLAK; COP, 2021).

A TD na Indústria 4.0 é complexa e exige muitos recursos, tornando uma diretriz de digitalização estratégica vital para o sucesso das pequenas e médias empresas na transição para a Indústria 4.0 (GHOBAKHLOO; IRANMANESH, 2021).

Selim (2021) aponta que a TD usa tecnologias de informação em muitas ocupações devido ao acesso rápido às informações para economizar custos e tempo excessivos. É possível ver benefícios em muitas atividades diárias, especialmente no mundo dos negócios. O quadro 1 apresenta uma lista de benefícios da TD.

Quadro 1 - Benefícios do uso de TD

<b>Tempo</b>	O fluxo fornecido com o arquivo Excel criado ou o existente software permite concluir o processo de trabalho em um tempo muito mais curto.
<b>Produtividade</b>	Com as ferramentas tecnológicas exigidas pela TD, o fluxo de trabalho é concluído muito mais rapidamente e a produtividade aumenta.
<b>Redução de despesas</b>	A transformação digital traz vantagens que vão aumentar rentabilidade operacional, graças ao "mais trabalho precisa de menos mão de obra" situação
<b>Redução de erros</b>	Com o software e automação robótica trazida pelo processo de TD, o trabalhos programados podem ser concluídos sem toque humano e sem erros
<b>Sustentabilidade</b>	Os sistemas de automação mecânica garantem ao negócio continuidade do processo, evitando interrupções que possam surgir de pessoal
<b>Consistencia</b>	Exposição mínima a quaisquer erros, problemas ou interrupções que possam surgir garante um processo de negócios constante
<b>Automação</b>	vantagens com relatórios, lembretes e gerenciamento de processos de negócios, que são fornecidos para executar mais trabalho produtivo colocando seu esforço nele
<b>Análise instantânea</b>	Graças à digitalização, é possível obter rapidamente informações sobre seus processos de negócios e analisar seu negócio
<b>Gestão de processos efetiva</b>	A gestão eficaz depende dados mensuráveis e tangíveis. É possível acessar facilmente os dados e avaliá-los periodicamente, graças às ferramentas digitais
<b>Inteligência de negócios alimentada por inteligência artificial</b>	Graças à IA é possível aplicar técnicas de análise de dados para fazer previsões e decidir sobre o novo caso de negócios se você tiver dados suficientes
<b>Aplicação da transformação digital nos negócios</b>	É muito mais rápido e fácil processar os dados que que transitam no ambiente digital.

Fonte: Autor (2024) adaptado de Selim (2021)

Selim (2021) aponta que a IA e o ML são impulsionados pela inovação em todos os setores e áreas funcionais.

Pode-se verificar, portanto, que a TD representa uma mudança significativa para o desempenho organizacional, e que ela incorpora tecnologias transformadoras como, por exemplo, *ML*.

Assim, estratégias de TD visam a transformação de produtos e processos na organização. Com foco explícito em atividades e processos digitais em uma organização a TD fornece a base necessária para adotar tecnologias avançadas que, por exemplo, no contexto deste estudo é da criação de uma ferramenta para apoiar no processo da definição de SLA.

Contextualizados os conceitos de transformação digital, fundamentais para atingir o objetivo do estudo, nota-se também a evolução da análise de grandes volumes de dados ou o

*Big Data*, sendo de suma importância na definição de estratégias de negócios e centrais de atendimento.

#### 1.4 *Big Data*

*Big Data* é frequentemente reconhecido como um termo usual para uma análise de dados mais inteligente e perspicaz, mas é mais do que isso, trata-se de novas fontes de dados desafiadoras que ajudam a entender os negócios em um nível mais granular, criando novos produtos ou serviços e respondendo às mudanças à medida que ocorrem. Como se vive em um mundo que constantemente produz e consome dados, é uma prioridade entender o valor que pode ser extraído deles (JUNIOR, 2021 *apud* DAVENPORT, 2012).

O valor potencial do *Big Data* segundo Gandomi e Haider (2015) é desbloqueado apenas quando aproveitado para conduzir a tomada de decisões baseada em evidências.

De acordo com Kostakis e Kargas (2021), as empresas foram e são inundadas com informações por meio de processamento de dados. A era digital levou as empresas a encontrar uma estratégia para se transformar a fim de superar as mudanças do mercado, competir com sucesso e ganhar uma posição de vantagem competitiva.

Segundo Lengauer (2020), estamos aprendendo com os dados há muito tempo. O processo geral é sempre o mesmo: o primeiro passo é coletar dados da forma mais sistemática possível. Isso pode acontecer apenas observando o sistema a ser investigado, ou também em um experimento controlado, que sistematicamente cria condições de contorno para as observações e realiza intervenções direcionadas.

Bello-Orgaz, Jung e Camacho (2016) complementam afirmando que *Big Data* é um fenômeno popular que visa fornecer uma alternativa às soluções tradicionais baseadas em bancos de dados e análise de dados. *Big Data* não é apenas sobre armazenamento ou acesso a dados. São soluções que possibilitam analisar os dados, a fim de dar sentido a eles e explorar o seu valor. O *Big Data* é uma fonte nova, porém poderosa, para potencialidades econômicas e valor social utilizada para obter uma vantagem competitiva em pé de igualdade com a organização de bens de capital e talento humano (GROVER *et al.*, 2018).

De acordo com Lee (2020) *Big Data* representa um novo paradigma tecnológico para dados que são gerados em alta velocidade, alto volume e com alta variedade. Ele é visto como um divisor de águas capaz de revolucionar a maneira como as empresas operam.

A combinação de tecnologias de *Big Data* e algoritmos tradicionais de ML gerou novos e interessantes desafios em outras áreas como mídia social e redes sociais (BELLO-ORGAZ; JUNG; CAMACHO, 2016).

Pode-se verificar pelo Quadro 2 no estudo Gupta e Rani (2019) a representação de definições que convergem com as definições previamente expostas.

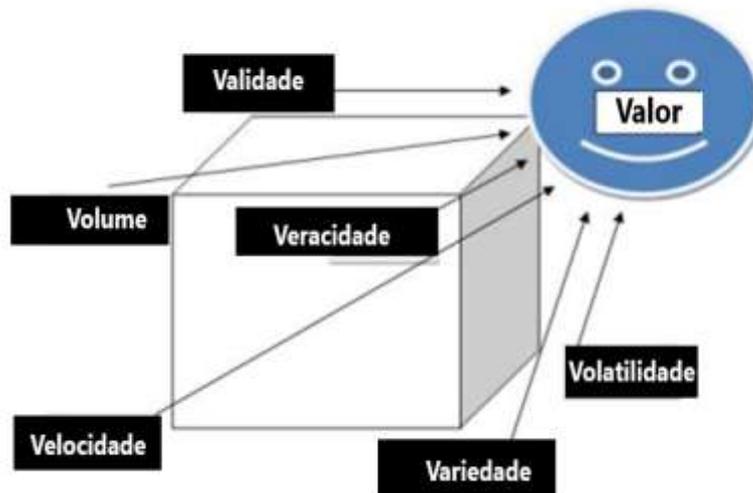
Quadro 2 - Definições de *Big Data*

<b>Autor</b>	<b>Ano</b>	<b>Definição</b>
McKinsey Global Institute	2011	<i>Big Data</i> é um conjunto de dados cujo tamanho está além do capacidade de ferramentas típicas de <i>software</i> de banco de dados para capturar, armazenar, gerenciar e analisar'
Gartner	2012	<i>Big Data</i> é alto volume, alta velocidade e/ou alta variedade ativos de informação que exigem custo-benefício, inovação formas de processamento de informações que permitem insight, tomada de decisão e automação de processos
Mayer-Schonberger e Cukier	2013	<i>Big Data</i> refere-se a coisas que alguém pode fazer em grande escala que não pode ser feito em um menor, para extrair novos insights ou criar formas de valor, de forma a mudar os mercados, organizações, a relação entre cidadãos e governos e muito mais
National Institute of Standards and Technology	2015	<i>Big Data</i> consiste em extensos conjuntos de dados - principalmente no características de volume, variedade, velocidade e/ou variabilidade - que exigem uma arquitetura escalável para eficiência armazenamento, manipulação e análise
International Data Corporation (IDC)	2017	<i>Big Data</i> é descrito como uma nova geração de software e arquiteturas projetadas para extrair valor de volumes muito grandes de uma ampla variedade de dados permitindo captura, descoberta e/ou análise'.

Fonte: Autor adaptado de GUPTA e RANI (2019)

Considerando a quantidade massiva de dados gerados pelas empresas Cavlak e Cop (2021), apontam que as empresas estão usando informações específicas, bem como *insights* sobre os clientes e seu comportamento a partir dos dados que coletam para se transformarem e, ao mesmo tempo, mudarem suas estratégias.

Compreender e discutir os V's abrirá portas para encontrar o verdadeiro valor do *Big Data* (KHAN; UDDIN; GUPTA, 2014) e para tanto trazem a Figura 3 **Erro! Fonte de referência não encontrada.**

Figura 3 - Os 7 V's do *Big Data*

Fonte: Autor adaptado de Khan, Uddin e Gupta (2014)

Para facilitar o entendimento do *Big Data* temos a representação de cada V no **Erro! Fonte de referência não encontrada.**, conforme o Quadro 3.

Quadro 3 - Representação de cada V

Volume	Refere-se ao tamanho dos dados que estão sendo criados de todas as fontes, incluindo texto, áudio, vídeo, social redes, estudos de pesquisa, dados médicos, imagens espaciais etc.
Velocidade	Com a alta velocidade de entrada de dados, as empresas têm de estar preparadas com tecnologia e mecanismos de banco de dados para processá-los.
Variedade	Os dados aparecem em muitas formas áudio, vídeo, texto, imagens o que traz complexidade e dificulta a gestão dos dados.
Veracidade	Representando o quão certos estamos sobre os dados armazenados.
Validade	Refere-se a possibilidade de uso dos dados com em relação ao uso pretendido. Em outras palavras, os dados podem não tem problemas de veracidade, mas pode não ser válido se não for devidamente entendido.
Volatilidade	Falando em volatilidade de <i>big data</i> , podemos facilmente lembrar a política de retenção de dados estruturados que implementamos a cada dia em nossos negócios. Quando o período de retenção expirar, podemos facilmente destruí-lo.
Valor	É considerado um V especial por um motivo. Ao contrário de outros V's este representa o desejado resultado do processamento de <i>big data</i> .

Fonte: Autor adaptado de Khan, Uddin e Gupta (2014)

Song e Zhu (2016) afirmam que estamos começando a ver os impactos do *Big Data* em todos os aspectos de nossas vidas e da sociedade. E, portanto, precisamos de uma ciência que possa resolver problemas do *Big Data*. A ciência de dados é uma nova disciplina emergente

que foi nomeada para abordar os desafios que estamos enfrentando e enfrentaremos na era do *Big Data*.

Evidencia-se assim a enorme relevância da quantidade de dados gerados hoje, inclusive pelos *helpdesks*, assim como a importância tomada pelo devido uso de sistemas que proporcionam retirar valor desses dados o que fez emergir uma ciência que viabiliza a análise e estudo dos dados.

## 1.5 Ciência de dados

Ao longo da história da humanidade, a forma como os humanos transmitem inteligência de geração em geração mudou várias vezes. Começou verbalmente e por meio de manuscritos, seguiu por invenções patenteadas, documentos oficiais e privados e hoje em dia, por diferentes formas de adaptar e implementar o conhecimento adquirido por meio de dados. Seja no que diz respeito à inteligência humana, artificial ou mista, os dados podem fornecer respostas consistentes e significativas para enfrentar os desafios dos negócios de hoje (TITU; STANCIU, 2020).

De acordo com Demchenko *et al.*, (2016) ciência de dados é um campo emergente da ciência, que requer uma abordagem multidisciplinar e deve ser construída com um forte vínculo com o *Big Data* e emergentes tecnologias.

A ciência de dados divide definições e áreas de atuação com outros campos, que suportam e guiam a extração de informações e conhecimento a partir de dados estruturados e não estruturados (SANTOS, 2017).

Pierson (2017) aponta que a ciência de dados é a ciência computacional de extrair insights significativos de dados brutos e, em seguida, comunicá-los efetivamente para gerar valor.

Embora a ciência de dados possa ser um tópico novo para muitos, é uma habilidade que qualquer indivíduo que deseja permanecer relevante em seu campo de carreira e necessidades da indústria saber (PIERSON, 2017).

Ostblom e Timbers (2022) colaboram e definem ciência de dados como o estudo, desenvolvimento e prática de processos reproduzíveis e auditáveis para obter insights de dados.

Como um novo campo, muito tem sido escrito sobre o uso de algoritmos de ciência de dados que podem gerar resultados úteis. De fato, muitos no campo, como Chen, Chiang e Storey (2012), acreditam que a pesquisa em ciência de dados precisa continuar para desenvolver análises avançadas e, portanto, não surpreendentemente, a pesquisa em ciência de dados normalmente se concentra em melhorar modelos de dados e os algoritmos usados em projetos de ciência de dados (SALTZ; SHAMSHURIM; CONNORS, 2013).

Priestley e Mcgrath (2019) apontam esta como uma disciplina que pode ser vista através das lentes de um novo modo para produção de conhecimento e caracteriza-se pela colaboração transdisciplinar com o setor privado de maior responsabilidade. As lições dessa evolução podem gerar produção de conhecimento em outras disciplinas acadêmicas tradicionais, bem como gerar práticas de gestão do conhecimento estabelecidas lidando com os desafios emergentes do *Big Data*.

A ciência de dados surge neste contexto, como ferramenta capaz de gerenciar um volume crescente de informação, fundamental no apoio à gestão (GUEDES JÚNIOR, 2021).

A evolução na transmissão de conhecimento ao longo da história foi abordada e o estudo definiu a ciência de dados apresentando a mesma como tópico central, pois por meio dela é possível extrair insights significativos de dados brutos. Com seu forte vínculo com Big Data e tecnologias emergentes, como inteligência artificial, a ciência de dados busca apresentar insights por meio de processos reproduzíveis e auditáveis.

A definição de ciência de dados traz consigo a definição de uma ferramenta promissora, multidisciplinar, necessária e que possibilita a descoberta de informação útil para o desenvolvimento de processos reproduzíveis e auditáveis que fornecem insights para a tomada de decisão mais assertiva na adequação e uso dos recursos de *helpdesk*. Nesse contexto faz-se necessário o uso da inteligência artificial, uma das tecnologias emergentes que é apresentada no próximo tópico.

## **1.6 Inteligência artificial**

Não há na literatura uma definição de consenso a respeito de IA. Esta indefinição reside nas diversas dimensões em que este tema pode ser abordado (MODESTI, 2020).

O desenvolvimento da IA tem experimentado vários ciclos de avanço desde seu início em 1943. Atualmente, as principais forças motrizes de avanço da AI são dados, processadores e algoritmos de aprendizado (MODESTI, 2020 apud XUEJIAO; XIAOFENG; YANG, 2013).

Um atributo que está diretamente vinculado ao que a IA visa alcançar é a criação de sistemas que possam entender seu ambiente e, conseqüentemente, tomar medidas para aumentar as chances de sucesso (MODESTI 2020 *apud* DOPICO, 2016).

De acordo com Mendonça e Dantas, (2020) TD, *Big Data*, inteligência artificial e análise de dados são provavelmente os assuntos computacionais mais citados nas notícias em todo o mundo.

Kaynak, (2021) aponta que durante as últimas duas décadas, profundas mudanças tecnológicas ocorreram ao nosso redor. E complementa, o desenvolvimento mais recente é a integração da IA à transformação digital onde está é tida como o principal habilitador e facilitador da criação de valor. Espera-se que as aplicações de IA realmente transformem nosso mundo de forma a impactar todas as facetas da sociedade, economia, vida, trabalho, saúde e tecnologia.

Malik, Chaudhary e Srivastava, (2022) apontam que a TD é a aquisição da ferramenta digital, técnicas, abordagens, mecanismo etc. para a transformação do negócio, aplicativos, serviços e atualização do processo manual para a automação. E complementam apontando que, devido ao alto impacto da inteligência artificial avançada, aliada a técnicas de ML e análise de dados, proporciona melhor aproveitamento do lucro ocasionado pela TD.

De acordo com Elrefai *et al.*, (2021) a IA e o ML estão transformando o negócio hoje em dia, pois possuem uma ampla gama de usos em todos os aspectos de qualquer organização.

Ainda que a definição de IA não tenha atingido sua plenitude, refletindo a multidimensionalidade desse campo. A importância da IA é inquestionável, e seu desenvolvimento tem sido impulsionado por avanços em dados, processadores e algoritmos de aprendizado. A crescente necessidade de informações em tempo real, juntamente com a transformação digital, Big Data e análise de dados, torna a IA um tema central nas notícias e nas discussões em todo o mundo. Com aplicações abrangentes em todas as áreas organizacionais a IA e o ML promovem a eficiência e a automação o que inclui as centrais de atendimento. Portanto, a IA está moldando o presente e o futuro, oferecendo oportunidades significativas, mas também desafios importantes.

No próximo tópico a discussão será levada a uma das áreas de uso da IA que é de interesse deste trabalho, pois vai ao encontro com o objetivo, uma vez que este tem em no processo proposto, o estudo de viabilidade de usar predição de dados.

### 1.7 *Machine learning*

Berral *et al.*, (2010) trazem que as áreas de ML e Mineração de Dados são intimamente relacionadas e preocupam-se com a obtenção de conhecimento a partir de dados, o que normalmente envolve a criação de modelos ou a descoberta de padrões em exemplos do passado de um comportamento do sistema com o mínimo de intervenção especializada possível. Afirmam que o papel dos métodos de ML é fornecer previsões de valores que são desconhecidos quando eles são gerados.

Segundo Hurwitz e Kirsh (2018), ML tornou-se um dos tópicos mais importantes nas organizações de desenvolvimento que procuram maneiras inovadoras de alavancar ativos de dados para ajudar a empresa a obter um novo nível de entendimento. Com os modelos apropriados de ML, as organizações têm a capacidade de prever continuamente as mudanças nos negócios, para que possam ter uma previsão do que virá a seguir.

De acordo com Hurwitz e Kirsh (2018), como os dados são constantemente adicionados, os modelos de ML garantem que a solução seja atualizada constantemente. O valor é claro e evidente: se você usar as fontes de dados mais apropriadas e em constante mudança no contexto de ML, terá a oportunidade de prever o futuro.

Preez e Oosthuizen (2019) trazem que ML está se tornando um conceito cada vez mais popular no mundo moderno, pois seu objetivo mais comum é otimizar sistemas, permitindo um uso mais inteligente de produtos e serviços. Na indústria de manufatura o ML pode levar a economia de custos, economia de tempo, aumento da qualidade e redução de desperdício. Concluem o trabalho apontando que no campo da fabricação sustentável, um crescente nível de ML é usado para enfrentar os crescentes requisitos de produção.

Em seu trabalho Musumeci *et al.*, (2019) trazem importantes conceitos sobre ML e apontam que os algoritmos visam extrair conhecimento dos dados, com base em algumas entradas de caracterização, muitas vezes referidas como atributos (*features*). Dependendo dos dados disponíveis e no objetivo do modelo a ser desenvolvido, técnicas de ML podem ser classificadas em alto nível nas seguintes categorias:

Algoritmos de ML supervisionados são fornecidos com entrada rotulada dados, ou seja, há um conjunto de amostras de dados de treinamento históricos contendo os valores de entrada (recursos) e os saída correspondente, ou seja, os rótulos. Tais rótulos podem ser valores numéricos em uma faixa contínua ou valores discretos/categóricos. Nestes dois casos supervisionados a aprendizagem toma a forma de uma regressão ou classificação problema, respectivamente. Em algoritmos de ML não supervisionados, os dados não são rotulados sendo o objetivo identificar se há semelhanças entre os dados (este é geralmente referido como *clustering*) ou se houver notáveis exceções no conjunto de dados (isso geralmente é referido como "detecção de anomalia" (MUSUMECI *et al.*, 2019).

Algoritmos semi-supervisionados são híbridos e englobam as duas categorias anteriores. Auxiliam na resolução de problemas onde apenas alguns pontos de dados são rotulados e a maioria deles não tem rótulo (MUSUMECI *et al.*, 2019).

Aprendizado por reforço (RL) é outra área do aprendizado em que um agente aprende como se comportar de maneira otimizada (ou seja, como maximizar a recompensa obtida ao longo de um determinado horizonte de tempo) interagindo com o ambiente, recebendo um feedback após cada ação e atualizando de acordo seu conhecimento (MUSUMECI *et al.*, 2019)

Em particular, no caso de aprendizagem supervisionada (ver a primeira linha da

Figura 5, a fase de treinamento inclui as seguintes etapas:

Dados brutos de treinamento são pré-processados para extrair e selecionar recursos contendo informações úteis para a tarefa de regressão/classificação, ou seja, mostrando correlação estatística com o valor/classe de saída.

Um algoritmo de aprendizado adequado é selecionado. Vários algoritmos de aprendizagem, com características diferentes em termos de precisão alcançável, escalabilidade e esforço computacional, podem ser propostos.

O algoritmo escolhido aprende a regressão/classificação modelo, ou seja, um mapeamento entre o espaço de recursos e as saídas associadas.

Para avaliar e melhorar as propriedades de generalização do algoritmo desenvolvido na fase de treinamento, o algoritmo ML pode ser aplicado sobre um ou mais conjuntos de dados de validação para quais rótulos também são conhecidos, a fim de ajustar o parâmetro do algoritmo para evitar *overfitting*.

Figura 4 - Modelo de construção e teste de ML



Fonte: Autor adaptado de Musumeci *et al.*, (2019)

De acordo com Musumeci *et al.*, (2019) assim que concluída a fase de treinamento, o algoritmo de ML pode ser usado em um conjunto de dados de teste contendo novas instâncias caracterizado pelo mesmo tipo de características do conjunto de treinamento (ver segunda linha da

Figura 5 e para o qual o correspondente saídas são conhecidas (ou seja, elas representam a verdade), mas não são usados para realizar a previsão. Em geral, no teste fase, são realizadas as seguintes etapas:

- Os dados são pré-processados para extração e seleção de recursos.
- O modelo aprendido é aplicado no conjunto de dados de teste.
- As saídas fornecidas pelo algoritmo são elaboradas para serem visualizadas e/ou validadas, comparando a saída do modelo com a verdade fundamental. Para fazer essa avaliação, um amplo conjunto de métricas é usado.

Por outro lado, no caso do aprendizado não supervisionado, o fase de treinamento é normalmente ignorada e apenas a fase de teste é desempenhado. No entanto, vale notar em alguns casos, por exemplo, quando realizando a detecção de anomalias, uma fase de

treinamento pode estar presente também em algoritmos não supervisionados (MUSUMECI *et al.*, (2019).

Para o fechamento deste tópico temos que embora os algoritmos de ML existam desde a década de 1950, seu impacto inicial foi insignificante. Bilalli *et al.*, (2019) trazem que com o aumento da disponibilidade de dados e do poder computacional, as ferramentas e algoritmos de ML estão fazendo avanços em diversas áreas. Inclusive seu sucesso levantou a necessidade de generalizar o uso do ML, ou seja, envolver até mesmo usuários não especialistas e sem proficiência em estatística e ML para realizar análises de dados o que torna o processo de análise de dados desafiador.

Essas etapas serão evidenciadas e discutidas nos subtópicos a seguir.

### 1.7.1 *Limpeza dos dados*

Na condução de uma análise bibliométrica sobre os métodos existentes de limpeza de dados Iwata e Ito (2023) trazem que a limpeza de dados é uma etapa fundamental no processo de análise de dados, que consiste em identificar e corrigir erros, inconsistências e anomalias presentes nos conjuntos de dados. Essa etapa é necessária para garantir a confiabilidade e a qualidade dos resultados obtidos a partir dos dados. Iwata e Ito (2023) *apud* Johnson *et al.* (2018), também trazem que a limpeza de dados desempenha um papel importante na preparação dos dados para análise, uma vez que dados sujos podem levar a conclusões imprecisas e enviesadas.

Denotando a importância da atividade por meio do esforço despendido para realizá-la, Rezig *et al.*, (2018) trazem que os cientistas de dados gastam mais de 80% de seu tempo (1) ajustando modelos de ML e (2) iterando entre a limpeza de dados e a execução do modelo de ML. Embora existam esforços para apoiar o primeiro requisito, atualmente não existe um sistema de fluxo de trabalho integrado que combine a limpeza de dados e o desenvolvimento de aprendizagem automática.

Com o objetivo de viabilizar uma simples visão geral dos métodos de detecção de *outliers* e explicar a necessidade de limpeza de dados na área de consumo de energia Jurg *et al.*, (2020) apontam que as técnicas estatísticas e a IA estão se tornando muito mais uma necessidade em um mundo acelerado, em vez de apenas um caso de uso teórico, corroborando, portanto, com o pensamento de Bhattacharya (2019). Complementam afirmando que para

satisfazer a necessidade de obter insights dos negócios o processo começa com a coleta e limpeza dos dados.

De acordo com Ilyas e Rekatsinas (2022) a grande implantação de sistemas de ML nos negócios, na ciência, no meio ambiente e em diversas outras áreas começou a ser verificada e foi percebida a forte dependência da qualidade dos dados de entrada desses modelos de ML para obter previsões ou insights confiáveis.

Em sua proposta de modelo para treinamento de ML com limpeza de dados Krishnan *et al.*, (2016) trazem que os bancos de dados podem ser corrompidos com vários erros, valores ausentes, incorretos ou inconsistentes. Cada vez mais, os pipelines modernos de análise de dados envolvem ML, e os efeitos dos dados sujos podem ser difíceis de depurar. Dados sujos são frequentemente escassos e soluções de amostragem ingênuas não são adequadas para modelos de alta dimensão.

Reforçando importância da limpeza de dados Tãm *et al.*, (2016) trazem que o ML geralmente requer uma grande quantidade de dados de treinamento, e esses dados muitas vezes obtidos de várias fontes costumam ser de baixa qualidade, como um grande número de problemas. Tãm *et al.*, (2016) complementam apontando duas técnicas utilizadas no ML: definir regras de limpeza manuais e usar amostras para ML ou métodos estatísticos.

Temos, portanto, que o ML geralmente requer uma grande quantidade de dados de treinamento, e esses dados que podem ser obtidos de várias fontes costumam ser de baixa qualidade. Os pesquisadores que usam ML geralmente aplicam algumas técnicas de limpeza de dados para limpar dados corrompido que podem ser a definição de regras de limpeza manuais ou usar amostras propostas por métodos estatísticos.

### 1.7.2 Análise exploratória

Os avanços na ciência de dados e no *Big Data* vem possibilitando novas abordagens na análise de dados que podem ser grandes aliadas na extração de informações que até agora eram ocultas. dos dados de operação da planta. Diante da afirmação Bezerra *et al.*, (2019) propõem a utilização de Análise Exploratória de Dados (*EDA*) como uma abordagem promissora baseada em dados para pavimentar alarmes industriais e análise de eventos.

De acordo com Langer e Meisen (2019) na análise exploratória de dados, conhecimento e experiência do domínio desempenham um papel central para extrair informações dos dados e

para obter provas e conhecimento. No entanto, especialistas raramente são as mesmas pessoas que realizam as análises de dados.

Barczewski *et al.*, (2020) trazem que a análise exploratória de dados é um processo iterativo aberto, onde o objetivo é descobrir novos insights. Muito de o trabalho para caracterizar esta exploração decorre de pesquisa qualitativa, resultando em descobertas ricas, taxonomias de tarefas e modelos conceituais.

Para Daele e Janssenswillen, (2023) as melhores práticas em (ensino) de alfabetização em dados, especificamente *exploratory data analysis (EDA)*, permanecem uma área de conhecimento tácito até este dia. No entanto, com o aumento da quantidade de dados e a sua importância nas organizações, a análise exploratória de dados está se tornando uma competência muito necessária. Descrevem um experimento empírico que foi usado para examinar as etapas executadas durante uma pesquisa exploratória de dados, análise e a ordem em que essas ações foram tomadas.

Podemos entender, portanto, avanços na ciência de dados e no *Big Data* o que viabiliza novas abordagens na análise de dados para extrair informações que anteriormente estavam ocultas nos dados. O uso de EDA é uma abordagem promissora e juntamente com o conhecimento e a experiência do domínio desempenham um papel central. Conceituada a análise exploratória temos o direcionamento para o próximo tópico.

### 1.7.3 Engenharia de atributos

De acordo com Davis e Foo (2016) a geração de recursos de entrada discriminativos é um requisito fundamental para obter classificadores altamente precisos. O processo de engenharia de atributos inicia a partir de dados brutos pode exigir esforço manual significativo.

Chen *et al.*, (2019) destacam a engenharia de atributos para o desenvolvimento de modelos eficazes de ML. Tradicionalmente, é realizada manualmente, o que requer muito conhecimento de domínio e é demorada. Nos últimos anos, muitos métodos automatizados de engenharia de atributos foram propostos com a finalidade de melhorar a precisão dos modelos de aprendizado.

Contribuindo com a definição Wang *et al.*, (2020) trazem que nos últimos anos, o ML desenvolveu rapidamente e tem sido amplamente aplicado em muitos campos. Muitos estudos mostraram que engenharia de atributos é a parte mais importante de ML e a parte mais criativa

da ciência de dados. No entanto, na etapa tradicional de engenharia de atributos, muitas vezes requer a participação de especialistas experientes no domínio e consome muito tempo.

Colaborando com a definição Siradjuddin *et al.*, (2021) trazem que a abordagem de engenharia de atributos determina recursos usando conhecimento especializado sobre a entrada de dados e o método de extração de recursos. Portanto, esta abordagem obtém recursos que representam a entrada para o problema de classificação. A abordagem de aprendizagem de recursos depende do processo de aprendizagem que mapeia a entrada de dados brutos e seu destino. Os recursos obtidos são a representação da entrada de dados com base em seu alvo.

Criar características ou variáveis que sejam eficazes na diferenciação entre diferentes classes ou categorias em um problema de ML ou reconhecimento de padrões é de suma importância. Temos, portanto, o reconhecimento da engenharia de atributos como uma parte fundamental e criativa da ML, sendo determinada por abordagens que envolvem conhecimento especializado e extração de recursos para representar efetivamente os dados de entrada no contexto da classificação.

#### 1.7.4 Pré-processamento

Carvalho *et al.*, (2011) apontam que apesar de algoritmos de ML serem frequentemente adotados para extrair conhecimento de conjuntos de dados, seu desempenho é geralmente afetado pelo estado dos dados que podem estar limpos ou conter algum ruído com valores incorretos, inconsistentes, duplicados ou ausentes e técnicas de pré-processamento de dados são frequentemente utilizadas para melhorar a qualidade dos dados por meio da eliminação ou redução dos problemas citados.

De acordo com Guimarães *et al.*, (2019) uma das etapas fundamentais do processo de classificação de textos é o pré-processamento, na qual um conjunto de ações transformadoras é aplicado, para garantir que a informação em formato não estruturado (texto) se torne passível de compreensão pela máquina.

O pré-processamento é um passo importante para selecionar dados de boa qualidade para utilização em operações de mineração. Mesmo assim, apesar de a seleção de atributos poder ser muito benéfica para pré-selecionar dados representativos visando melhorar o desempenho da previsão final, não se sabe que método de seleção é o melhor (HORTA *et al.*, 2014).

Com uma visão mais abrangente da atividade Santos *et al.*, (2019) trazem que o pré-processamento é guiado pelos algoritmos que serão utilizados para o ajuste de modelos preditivos. De modo geral, sua aplicação está relacionada às seguintes atividades: (1) transformação de variáveis quantitativas (via padronização ou normalização); (2) redução de dimensionalidade do conjunto de dados (exclusão de preditores altamente correlacionados ou utilização de análise de componentes principais); (3) exclusão de variáveis/observações com dados faltantes ou utilização de técnicas de imputação (média, mediana ou valor mais frequente no caso de preditores numéricos ou categoria adicional, representativa dos indivíduos sem informação, no caso de preditores categóricos); (4) organização de variáveis qualitativas (decomposição das variáveis categóricas em um conjunto de variáveis indicadoras que serão utilizadas como preditores)

O objetivo do pré-processamento é extrair, de textos escritos em linguagem não estruturada, uma representação estruturada e manipulável por algoritmos de classificação que identifique o subconjunto mais significativo para a padronização da coleção de textos. Basicamente, a seleção reduz o conjunto de palavras que representará o documento no processo de classificação. São exemplos das etapas de seleção a remoção de palavras irrelevantes, como preposições ou artigos, a formatação dos textos e o cálculo de relevância dos termos para identificar os mais significativos. Por fim, os textos são representados em um formato estruturado que preserve as principais características dos dados (GUIMARÃES *et al.*, 2019 *apud* REZENDE 2011).

Em seu trabalho Kewo *et al.*, (2019) apontam ter encontrado valores ausentes, redundância e inconsistência nos conjuntos de dados e, na maioria dos casos, o pré-processamento dos dados é inevitável. Essa etapa converte os dados em um conjunto de dados limpo e organizado para as etapas de modelagem subsequentes ou análises estatísticas.

Colaborando com a definição Bilalli *et al.*, (2019) trazem que o pré-processamento de dados é uma das etapas mais demoradas e relevantes em um processo de análise de dados (por exemplo, tarefa de classificação).

Trazendo a discussão para um contexto em que se faz necessária a análise de grandes quantidades de dados Bekar *et al.*, (2020) trazem que as atuais aplicações de manutenção preditiva no ambiente fabril são agora mais dependentes de algoritmos de ML baseados em dados, exigindo uma análise inteligente e eficaz de grandes quantidades de dados históricos e em tempo real provenientes de vários fluxos (sensores e sistemas de computador) disponíveis em vários equipamentos.

A relevância do pré-processamento de dados, conforme destacado por diversos autores, é incontestável no contexto da aplicação de algoritmos de ML em diversas áreas, desde a classificação de textos até a manutenção preditiva em ambientes fabris. Os desafios relacionados à qualidade dos dados, como valores ausentes, redundância e inconsistência, destacam a necessidade inevitável da limpeza para assegurar conjuntos de dados limpos e organizados. O pré-processamento não apenas possibilita a aplicação eficaz de algoritmos de ML, mas também permite a extração de informações relevantes a partir de grandes volumes de dados, seja no contexto de análises estatísticas, classificação de textos, ou em ambientes IoT. Portanto, sua execução cuidadosa e eficiente é fundamental para garantir a confiabilidade e eficácia dos modelos preditivos e análises resultantes.

Esgotando-se os tópicos deste referencial teórico é apresentada a seguir a metodologia utilizada para realizar o trabalho.

## 2. METODOLOGIA

No que diz respeito a métodos de pesquisa, podemos conceituá-los como um conjunto de passos reconhecidos pela comunidade acadêmica e utilizados pelos pesquisadores para a construção do conhecimento científico (DRESCH *et al.*, 2015 *apud* ANDERY *et al.*, 2004).

Na busca pelo rigor metodológico Dresch *et al.*, (2015) apontam ser necessário que o pesquisador defina, logo no início das suas atividades, qual será o método sendo fundamental a explicação dos motivos que conduziram a essa escolha.

O método de pesquisa *Design Science Research (DSR)* foi o escolhido para orientar esse trabalho de pesquisa, pois, dentre os objetivos há o de produzir e avaliar um processo que suporte a solução para um problema vivenciado por muitas organizações, neste caso, o não atendimento de um chamado dentro do prazo. Temos, de acordo com Dresch *et al.*, (2020), que as pesquisas realizadas sob o paradigma das ciências tradicionais resultam em estudos que se concentram em explicar, descrever, explorar ou predizer fenômenos e suas relações sendo que, essas podem apresentar limitações.

Embora a comparação entre as ciências tradicionais e a DSR seja necessária deve ficar claro que elas não se opõem. A diferença entre as ciências tradicionais e a DSR pode ser verificada no produto gerado, enquanto a DSR está orientada para a geração de conhecimento que possibilite a resolução de problemas a pesquisa tradicional está voltada para fundamentos como explorar, descrever, explicar (DRESCH *et al.*, 2020).

Gregor e Hevner (2013) colaboram e trazem que a pesquisa científica em DSR conquistou seu devido terreno como um importante e legítimo paradigma de pesquisa de sistemas. Afirmam que a *DSR* ainda não atingiu todo o seu impacto potencial no desenvolvimento e uso de sistemas de informação devido a lacunas na compreensão e aplicação dos conceitos de *DSR* e métodos.

Para Silva (2020), a pesquisa em DSR se concentra no desenvolvimento de artefatos envolvendo duas atividades principais para melhorar e entender o comportamento dos aspectos dos sistemas de informações: (1) a criação de novos conhecimentos por meio do design de artefatos novos ou inovadores (coisas ou processos) e (2) análise do uso e/ou desempenho do artefato com reflexão e abstração.

Para Dresch, Lacerda e Júnior (2020), DSR como método de pesquisa orientado à solução de problemas a *DSR* busca a partir do entendimento do problema, construir e avaliar artefatos que permitam transformar situações, alterando suas condições para estados melhores ou desejáveis.

De acordo com Cauchick-Miguel, (2007) uma pesquisa desenvolve-se ao longo de um processo que envolve inúmeras fases, desde a adequada formulação do problema até a satisfatória apresentação dos resultados, análise crítica e suas conclusões.

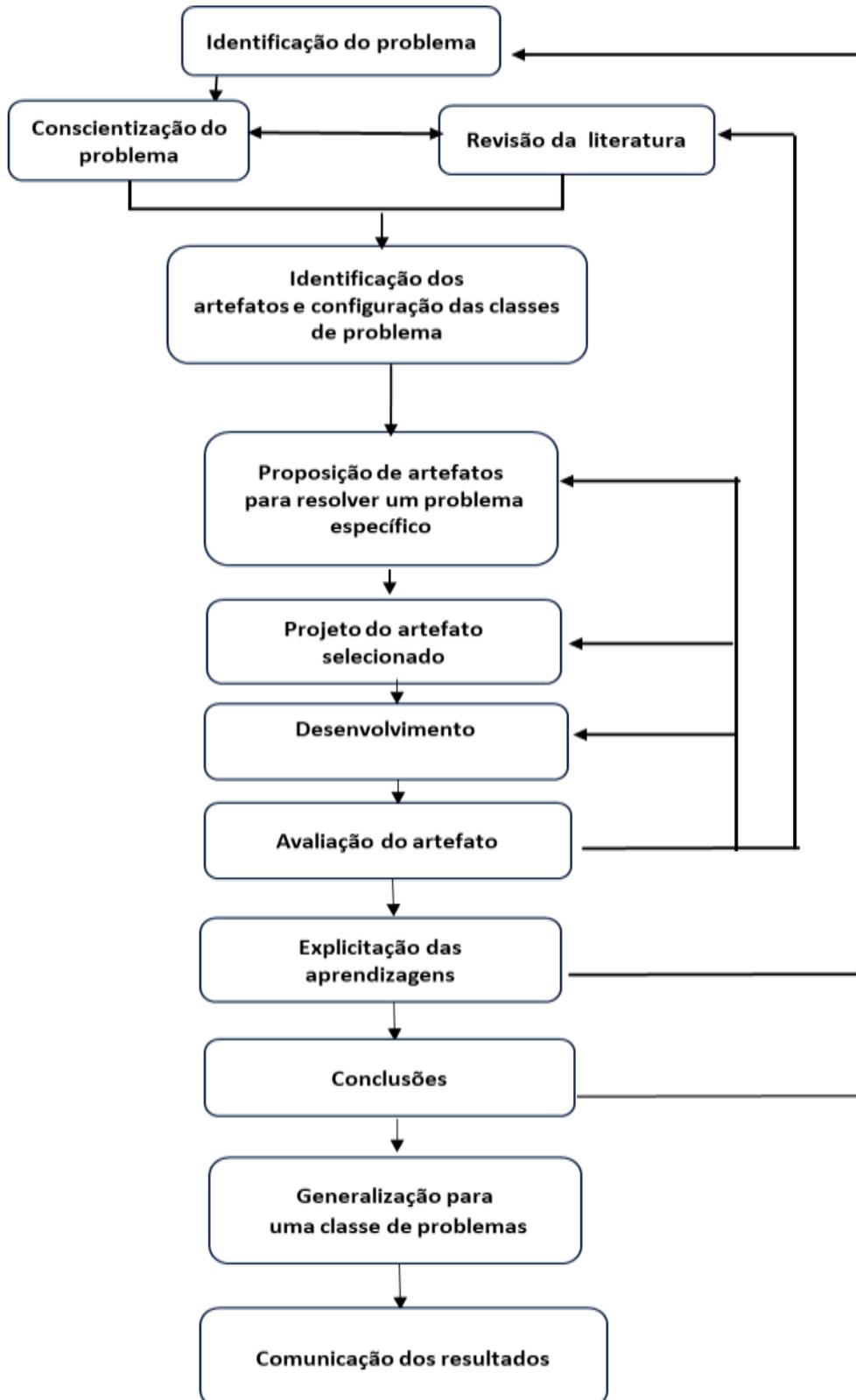
O processo de pesquisa é desenvolvido mediante o concurso do conhecimento disponível e a utilização cuidadosa de métodos, técnicas e outros procedimentos científicos (CAUCHIK-MIGUEL, 2007).

Neste estudo o procedimento de pesquisa é do tipo qualitativo que, de acordo com Gerhardt e Silveira, (2009) busca explicar o porquê das coisas, exprimindo o que convém ser feito, não quantificando os valores e as trocas simbólicas nem a submissão à prova de fatos, pois os dados analisados são não-métricos (suscitados e de interação) e se valem de diferentes abordagens, é também de cunho exploratório que, neste caso, segundo Vieira, (2002) visa proporcionar ao pesquisador uma maior familiaridade com o problema em estudo por meio de métodos bastante amplos e versáteis que compreendem: levantamentos em fontes secundárias (bibliográficas, documentais etc.), levantamentos de experiência, estudos de casos selecionados e observação informal.

Colaborando com a definição Garvey e Jones (2021) trazem que a pesquisa qualitativa parte da posição de que não existe uma realidade observável. Os pesquisadores que utilizam métodos qualitativos constroem descobertas indutivamente, desde dados brutos até uma compreensão conceitual. a rica descrição fornecida pelas análises pode permitir que a análise seja mais profunda.

Dresch *et al.*, (2015) apontam que além da forma de conhecimento gerado por DSR ser diferenciada o objetivo e conhecimentos gerados, por consequência, também são distintos e traz uma proposta para condução de pesquisas utilizando o DSR representada pela

Figura 5 - Etapas propostas para a condução de pesquisas utilizando DSR



Fonte: Autor (2024)

O método utilizado neste estudo inicia-se pela etapa de identificação do problema e prossegue pela sua conscientização, nesta etapa temos a interação direta com a revisão sistemática da literatura e traz como resultado a identificação dos estudos que de alguma forma

podem propor uma resolução do problema em questão. Como não foi possível identificar propostas de estudos que proporcionasse uma base para a solução foi iniciada a etapa do projeto. Nessa fase os pesquisadores devem transformar a proposta abstrata em um design concreto. Feedback contínuo e revisões são comuns, sendo necessário, por vezes, visitar a proposição do artefato para realizar ajustes com base no design detalhado, passando pelo desenvolvimento, fase altamente interativa e técnica que envolve a construção real do artefato conforme especificado na fase de design. Nesta fase os pesquisadores podem encontrar desafios técnicos e limitações que exigem ajustes no design o que cria um ciclo iterativo entre o projeto e o desenvolvimento e modificações são feitas conforme necessário para garantir que o artefato funcione conforme esperado.

Na avaliação o artefato é avaliado quanto à sua eficácia em resolver o problema específico. Esta fase pode envolver testes empíricos, simulações, ou feedback de usuários. Os resultados da avaliação podem levar a uma revisão das fases anteriores. A interação aqui é contínua até que o artefato atinja um nível satisfatório de desempenho.

Considerando o modelo proposto neste estudo a identificação do problema foi colocada na introdução onde se apresenta a necessidade de se manter a qualidade do atendimento de *helpdesk* e a melhoria de assertividade ao SLA em organizações e considerar a transformação digital para manter o GSTI, auxiliando as organizações nesse processo para melhorar a qualidade e a entrega de serviços e gerar mais valor. O sistema de *helpdesk* de TI é importante para grandes instituições que dependem fortemente de serviços e recursos de TI, pois simplifica o gerenciamento de tickets, automatiza tarefas diárias de TI e contribui para a definição e minimização de atividades nos processos de negócios. Temos, portanto, como notória a existência de dificuldade na definição do SLA que seja exequível.

Com o propósito de elevar não apenas a disponibilidade dos serviços oferecidos pela área de TI aos usuários, para as organizações, há também a possibilidade de melhorar a qualidade dos serviços e reduzir as reclamações dos clientes o que traz benefícios tanto para prestadores quanto tomadores dos serviços.

No esforço para compreender o problema, segunda etapa de acordo com o modelo de Dresch., *et al* (2015), foi entendido que o GSTI possibilita a avaliação do desempenho do atendimento com base em indicadores- existentes em uma base de dados que pode direcionar ações corretivas ou até mesmo preventivas. É evidente que há investimentos dedicados à melhoria da qualidade dos serviços com o objetivo de criar valor e que o sistema de *helpdesk*, é vital para as organizações, mas um processo que permita um maior conhecimento com base

nas informações disponíveis em base de dados pode ser uma ferramenta de apoio à melhoria da prestação de serviços inclusive usando recursos de ML.

Segundo Dresch., *et al* (2015), a importância de buscar conhecimento possibilita classificar os trabalhos e faz com que o pesquisador se depare com um estudo semelhante ou até mesmo ideal e pronto, e, sendo o caso, o pesquisador poderá continuar com a pesquisa à medida em que o novo estudo traga melhores soluções em comparação ao existente. Neste sentido foi feita uma revisão sistemática da literatura nas bases de pesquisa *Web of Science* e *Scopus*.

Em seguida foi proposto e construído um processo como solução para diagnóstico e melhoria do atendimento de *helpdesk* denominado PRODIMAH (Processo que prevê o uso de estatística descritiva e ML. O processo construído foi aplicado para a avaliação em uma empresa de capital 100% nacional, especializada em TI e Integradora de Soluções cujo principal foco é o fornecimento de soluções integradas em serviços de TI, softwares, aplicativos e sistemas de gestão empresarial, empregando as mais conceituadas plataformas globais para os setores de Agronegócio, Manufatura, Automotivo, Logística, Serviços e Varejo.

As fases de identificação dos estudos e configuração das classes de problemas, proposição, projeto e desenvolvimento do artefato bem como sua aplicação para avaliação são apresentadas no capítulo de resultados após o capítulo de revisão da literatura.

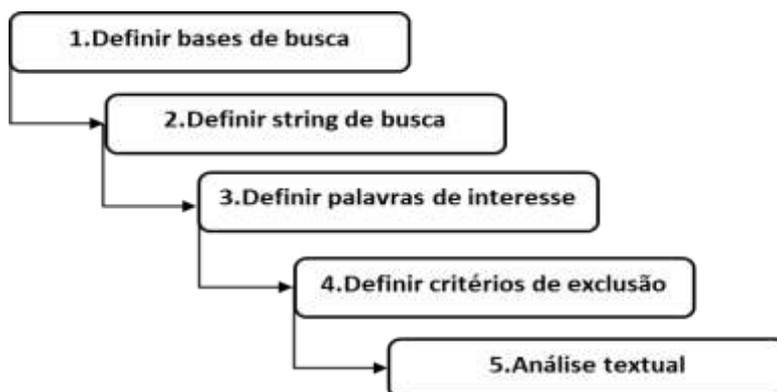
### 3. REVISÃO SISTEMÁTICA DA LITERATURA

Nesta etapa da proposta para condução de pesquisas que utilizam DSR, Dresch., (2015) apontam que é importante que o pesquisador faça uma consulta à bases de conhecimento por meio de uma revisão sistemática da literatura.

Para responder as questões “Quais as técnicas de ML têm sido usadas para a manutenção de SLA?” e “Para o que ML e seu uso na indústria de serviços de TI, especialmente no *helpdesk*. Foi feita uma revisão sistemática da literatura que, de acordo com Sharif *et al.*, (2019) tem sido considerada como um exercício essencial para avaliar a natureza do conhecimento produzido em um campo de pesquisa, suas lacunas e possíveis desenvolvimentos futuros com o objetivo de identificar as técnicas de ML utilizadas no suporte ao SLA. A revisão sistemática da literatura visou explorar a relação entre Inteligência Artificial, ML e SLA, ressaltando a importância do SLA como contrato que define os níveis de qualidade dos serviços que, se insatisfatórios, podem trazer consequências negativas tanto para tomadores quanto prestadores de serviço e como a IA pode ajudar a prevenir violações.

Temos, conforme Figura 6, as macro etapas para atingir o objetivo da revisão sistemática da literatura.

Figura 6 - Macro etapas da revisão sistemática da literatura



Fonte: Autor (2024)

Na primeira etapa foram definidas as bases *Web of Science* e *Scopus* por terem aderência ao perfil de artigos procurados.

Na sequência houve a definição das strings de busca conforme Quadro 4 com o uso de termos relacionados ao objetivo da pesquisa.

Quadro 4 - Strings de busca

Base de conhecimento	Texto da String
Web of Science	( <i>Helpdesk</i> OR "Help Desk" OR "IT Support" OR "Support Center" OR "IT incident resolution" OR "incident resolution" OR "Ticket automation" OR "Support ticket" OR "information technology services management") (Todos os campos) AND (PREDICT* OR "Machine Learning" OR "Artificial Intelligence" OR AI OR ML) (Todos os campos)
Scopus	( <i>Helpdesk</i> OR "Help Desk" OR "IT Support" OR "Support Center" OR "IT incident resolution" OR "incident resolution" OR "Ticket automation" OR "Support ticket" OR "information technology services management") (Todos os campos) AND (PREDICT* OR "Machine Learning" OR "Artificial Intelligence" OR AI OR ML) (Todos os campos)

Fonte: Autor (2024)

Com a string na base *Web of Science* em uma busca mais abrangente foram encontrados 1047 trabalhos que foram reduzidos a 1044 após a seleção de trabalhos em inglês. Em análise inicial verificou-se que muitos trabalhos com palavras chaves muito gerais como “*help*” (presente em 133 resumos de artigos), “*predict*” (presente em 588 resumos de artigos) e “*support*” (presente em 255 resumos de artigos) não tinham relação com o objetivo desta pesquisa. Portanto, a fim de filtrar melhor os resultados foram definidas palavras de interesse que foram encontradas no título, palavras-chave ou resumo o que possibilitou selecionar 44 trabalhos com pelo menos uma das palavras de interesse a fim de complementar a string de busca utilizada sendo elas: *Help*, *support*, *incident*, *ticket*, *predict*, *Machine Learning*, *Artificial Intel e desk*, consideradas palavras de interesse conforme Quadro 5.

Quadro 5 - Web of Science - Palavras de interesse

Palavra	Título		Palavras chave		Resumo	
	Total da busca	Artigos aderentes	Total da busca	Artigos aderentes	Total da busca	Artigos aderentes
Help	40	18	18	0	133	0
support	82	10	38	0	255	0
Incident	10	1	6	0	28	1
Ticket	15	1	9	1	27	0
Predict	100	0	39	1	588	0
Machine Learning	22	2	39	1	66	0
Artificial intelligence	8	0	14	0	28	0
Desk	47	7	20	1	2	0

Fonte: Autor (2024)

O quadro 5 apresenta as palavras-chave encontradas no título, nas palavras-chave e no resumo. À direita de cada uma destas colunas são apresentados os artigos desse total dessas colunas que realmente eram aderentes ao tema da pesquisa.

Com a string na base *Scopus* foram encontrados 1893 trabalhos. Verificou-se na base *Scopus* a mesma situação onde muitos trabalhos com palavras chaves muito gerais como “*help*” (presente em 320 resumos de artigos), “*predict*” (presente 1082 resumos de artigos) e “*support*” (presente em 201 resumos de artigos) não tinham relação com o objetivo desta pesquisa.com a mesma estratégia utilizada na base *Web of Science* foram selecionados 56 trabalhos em função de terem as palavras de interesse conforme Quadro 6

Quadro 6 - Scopus - Palavras de interesse

Palavra	Título		Palavras chave		Resumo	
	Total da busca	Artigos aderentes	Total da busca	Artigos aderentes	Total da busca	Artigos aderentes
Help	47	29	33	0	320	0
support	190	8	106	0	201	2
Incident	19	5	14	0	38	0
Ticket	32	1	19	3	63	0
Predict	202	0	116	0	1082	2
Machine Learning	79	0	138	1	259	0
Artificial intelligence	29	0	64	3	181	0
Desk	48	1	33	1	129	0

Fonte: Autor (2024)

O quadro 6 apresenta as palavras-chave encontradas no título, nas palavras-chave e no resumo. À direita de cada uma destas colunas também são apresentados os artigos desse total dessas colunas que realmente eram aderentes ao tema da pesquisa.

De 100 artigos previamente selecionados 27 foram identificados em ambas as bases e apenas os artigos da *Web of Science* foram considerados. Resultando, inicialmente, em 73 trabalhos para análise textual.

Foram excluídos 8 estudos fora do tema, 5 estudos foram excluídos por não serem artigos, 10 artigos foram desconsiderados por não terem DOI, ISBN ou URL e por fim 14 artigos não foram considerados haja vista a impossibilidade de acessá-los.

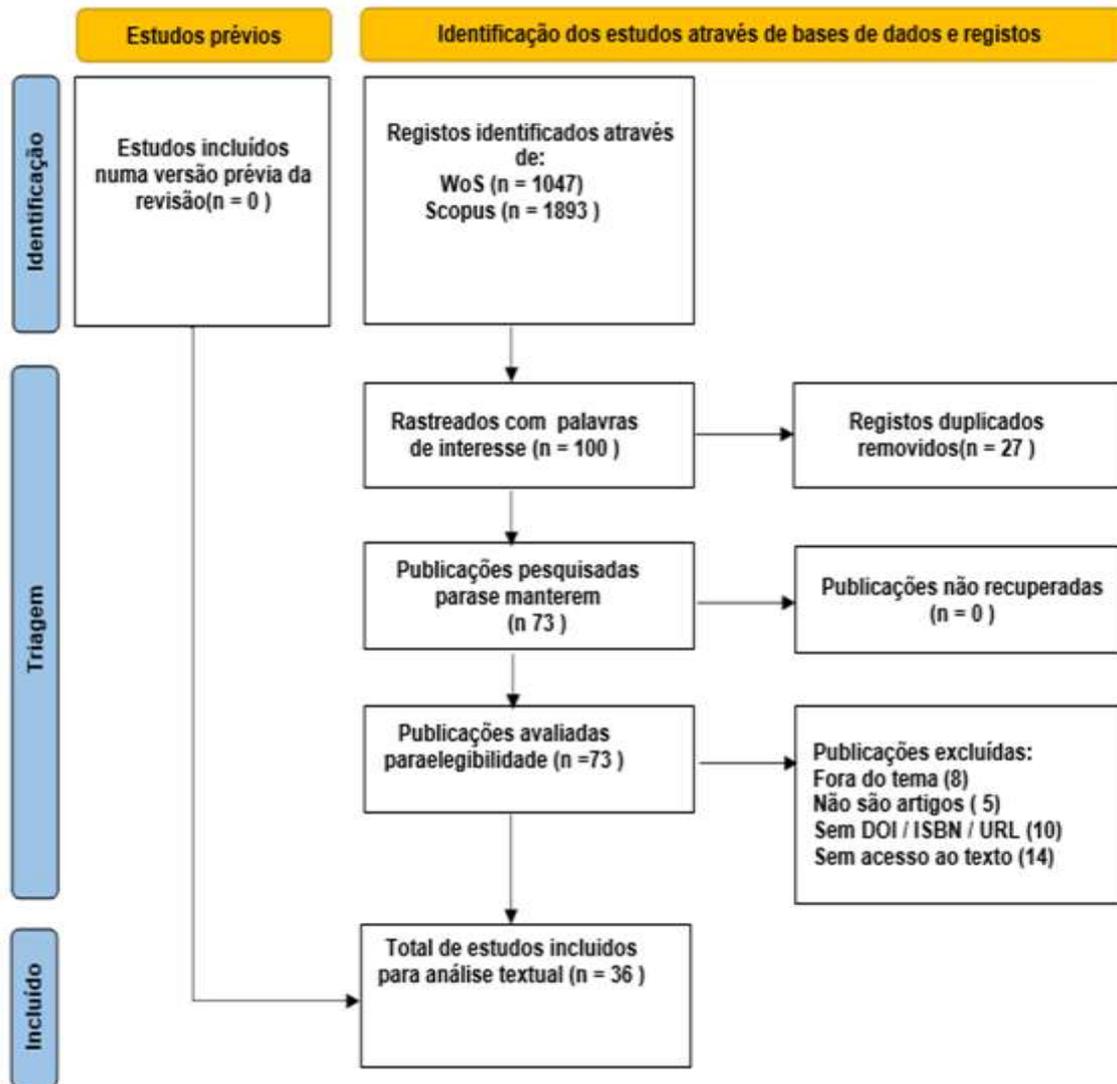
Os critérios de exclusão dos textos retornados na pesquisa após a exclusão das duplicidades foram:

- a. Fora do Tema da Pesquisa
- b. Idioma não dominado pelos autores

- c. Não se enquadra como artigo científico
- d. Sem DOI
- e. Sem acesso ao texto

Considerando os critérios de exclusão ficou, portanto, o montante de 36 artigos para análise textual conforme fluxo de atividades apresentado pela Figura 7.

Figura 7 - Fluxo para revisões sistemáticas atualizadas que incluíram apenas buscas em bases de dados e registros



Fonte: Autor (2024)

Com a análise dos artigos selecionados foi possível elaborar o

Quadro 7 com o ano, nome do artigo, classificando-os de acordo com a técnica utilizada assim como sua finalidade.

Quadro 7 -Análise dos artigos selecionados da revisão sistemática

Ano	Artigo	Técnica Utilizada	Para que foi utilizada
1993	Integrating Natural-Language Processing and Information-Retrieval In A Troubleshooting Help Desk	NLP	Entender as consultas em linguagem natural realizando processamentos como remoção de sufixos, eliminação de palavras de ruído e adição de sinônimos
1996	A Self-Improving <i>Helpdesk</i> Service System Using Case-Based Reasoning Techniques	CBR	Melhorar o atendimento ao cliente e a eficiência no gerenciamento de problemas em um ambiente de <i>helpdesk</i>
1997	Design Of an Intelligent Web-Based Help Desk System	CBR	Automatizar o <i>Helpdesk</i> oferecendo serviços World Wide Web
1999	Development And Utilization Of A Case-Based Help-Desk Support System in A Corporate Environment	CBR	Desenvolvimento do sistema de suporte de help desk baseado em casos HOMER
2001	A web-based intelligent fault diagnosis system for customer service support	CBR e Redes Neurais Artificiais (ANN)	Melhorar a eficiência e a qualidade do suporte ao cliente, reduzir os custos e acelerar os ciclos de serviço
2006	A Corpus-Based Approach to Help-Desk Response Generation	SVM	Prever respostas do <i>helpdesk</i> a partir de recursos nos e-mails dos clientes
2008	Predicting Service Request In Support Centers Based on Nonlinear Dynamics, Arma Modeling and Neural Networks	ARMA	Prever solicitações de serviço em <i>helpdesks</i>
2009	An Empirical Study Of Corpus-Based Response Automation Methods for An E-Mail-Based Help-Desk Domain	CBR	Automação respostas por e-mail do <i>helpdesk</i>
2009	Implementing An Online Help Desk System Based on Conversational Agent	NLP	Capacidade de fornecer atendimento ao cliente de forma muito mais econômica e interativa do que com métodos tradicionais. O sistema proposto demonstra uma nova forma de serviço eletrônico para uma organização aumentar a satisfação e retenção de seus clientes, levando a uma vantagem competitiva sobre outras organizações
2012	Hierarchical Online Problem Classification for IT Support Services	SVM	O foco principal é categorizar o problema que um usuário enfrenta ao reconhecer a especificidade real da causa raiz, aproveitando os dados de treinamento disponíveis, como monitoramento e registros nos sistemas
2016	Decision Trees for <i>Helpdesk</i> Advisor Graphs	Decision Tree	Facilitar o aconselhamento no trabalho de analistas juniores ou pessoal menos experiente sobre como abordar melhor relatórios de falhas de clientes de telecomunicações
2016	Emotion Recognition From <i>Helpdesk</i> Messages	SVM	Reconhecer 5 emoções diferentes dos usuários
2016	Towards Better Help Desk Planning: Predicting Incidents and Required Effort	Não cita a técnica utilizada mas faz referência a uma técnica estatística para agrupar dados ou objetos semelhantes em grupos	Desenvolver um modelo de custo para operações de <i>helpdesk</i>

Fonte: Autor (2024)

Quadro 7 -Análise dos artigos selecionados da revisão sistemática (continuação)

Ano	Artigo	Técnica Utilizada	Para que foi utilizada
2017	Understanding The Service Desk: Applied Forecasting and Analytics Approach	Modelo estatístico próprio	A partir de series temporais passadas estimar a quantidade de chamados a fim de viabilizar a distribuição correta de recursos
2018	A Comparative Study of Supervised and Unsupervised Classifiers Utilizing Extractive Text Summarization Techniques to Support Automated Customer Query Question-Answering	NLP	Extração semântica e de palavras-chave das consultas dos clientes
2018	A Data-Driven Approach To Predict an Individual Customer's Call Arrival In Multichannel Customer Support Centers	Logistic Regression	Prever a chegada de uma chamada de um cliente individual em Centro de Suporte ao Cliente
2018	Agent Assist: Automating Enterprise It Support Help Desks	NLP	Melhorar a eficiência da equipe de suporte de TI, fornecendo respostas procedimentais e Determinar a natureza
2018	Cognitive System to Achieve Human-Level Accuracy in Automated Assignment of <i>Helpdesk</i> Email Tickets	SVM	do problema mencionado em um tíquete de e-mail de entrada e, em seguida, encaminhá-lo automaticamente para um grupo (ou equipe) de resolução apropriado
2018	Customer Support Ticket Escalation Prediction Using Feature Engineering	CHAID, SVM, Logistic Regression	Prever o risco de escalonamento de chamados
2018	How Angry Are Your Customers? Sentiment Analysis of Support Tickets That Escalate	Sentiment analysis tools	Prever de forma confiável quando um ticket de suporte se torna um candidato para escalonamento com base na análise de sentimento das conversas dos analistas de suporte com os clientes
2019	Automated Dispatch Of <i>Helpdesk</i> Email Tickets: Pushing the Limits With Ai	KNN, SVM e Linear Regression	Atribuição e o encaminhamento de tíquetes
2019	Automated IT Service Desk Systems Using Machine Learning Techniques	SVM menciona a utilização de Naive Bayes, Logistic Regression, K-Nearest neighbor e SVM	Criar sistema de classificação de tickets
2019	Corporate It-Support <i>Helpdesk</i> Process Hybrid-Automation Solution with Machine Learning Approach	ANN - Artificial Neural Network	Fornecer soluções instantâneas em tempo real e categorização de e-mails.
2019	Itsm Automation - Using Machine Learning to Predict Incident Resolution Category	NLP, SVM, Logistic regression, KNN	Predizer a categoria de solução de um incidente
2020	A Hybrid Machine Learning Framework of Gradient Boosting Decision Tree and Sequence Model for Predicting Escalation in Customer Support	rede neural recorrente (RNN)	Prever se um determinado caso de suporte será escalado por um cliente no futuro próximo com base em casos de escalada anteriores
2020	Automated Assignment Of <i>Helpdesk</i> Email Tickets: An Artificial Intelligence Life-Cycle Case Study	SVM, KNN, Linear Regression	Solução abrangente e eficaz para a atribuição automatizada de tickets de e-mail de <i>helpdesk</i> , com a capacidade de se adaptar às necessidades em constante mudança do negócio e proporcionar economias significativas de recursos humanos.
2020	Reply Using Past Replies-A Deep Learning-Based E-Mail Client	NLP	Lidar com a sobrecarga de e-mails, utilizando modelos de processamento de linguagem natural e representação vetorial de texto para melhorar a eficácia na gestão de e-mails

Fonte: Autor (2024)

Quadro 7 -Análise dos artigos selecionados da revisão sistemática (continuação)

Ano	Artigo	Técnica Utilizada	Para que foi utilizada
2021	A Machine Learning Based Help Desk System for It Service Management	SVM, KNN e decision tree	Melhorar a eficiência e a satisfação do usuário por meio da automação
2021	Question Classification Framework For Helpdesk Ticketing Support System Using Machine Learning	Naïve Bayes (NB) and Support Vector Machine (SVM)	Automatizar o processo de classificação de perguntas aos usuários
2022	Automated Support Request Categorization Using Machine Learning	Decision Tree e ANN - Artificial Neural Network	Categorização automática de solicitações/tickets de suporte ao usuário
2022	Automation Of Incident Response and It Ticket Management by MI and Nlp Mechanisms	NLP	Automatizar a resposta a incidentes de usuários
2022	Improving The Prediction Resolution Time for Customer Support Ticket System	Random Forest (RF) regression, Neural Network (NN) and ADA boost.	Prever tempo de resolução de chamados
2022	Problem Classification for Tailored Helpdesk Auto-Replies	NLP	Adaptar o conteúdo da resposta automática ao problema do usuário, de modo a aumentar a relevância das informações incluídas.
2023	A Hybrid Continual Learning Approach for Efficient Hierarchical Classification of It Support Tickets In The Presence Of Class Overlap	SVM, Logistic Regression e Decision Tree	Classificar tickets de suporte ao cliente de acordo com o desejado critério
2023	Hyperparameter Black-Box Optimization to Improve The Automatic Classification Of Support Tickets	CNN e SVM	Categorização automática de solicitações/tickets de suporte ao usuário
2023	Key Factors In Achieving Service Level Agreements (SLA) For Information Technology (It) Incident Resolution	Logistic Regression and Neural Network (NN)	Analisar o impacto de vários fatores no cumprimento de acordos de nível de serviço (SLAs) para a resolução de incidentes

Fonte: Autor (2024)

A análise textual permitiu concluir que diversas técnicas são aplicadas concomitantemente. Sendo assim, foi feita a contagem das técnicas de acordo com sua utilização em cada respectivo caso gerando o Quadro 8.

Quadro 8 - Técnicas utilizadas e suas quantidades

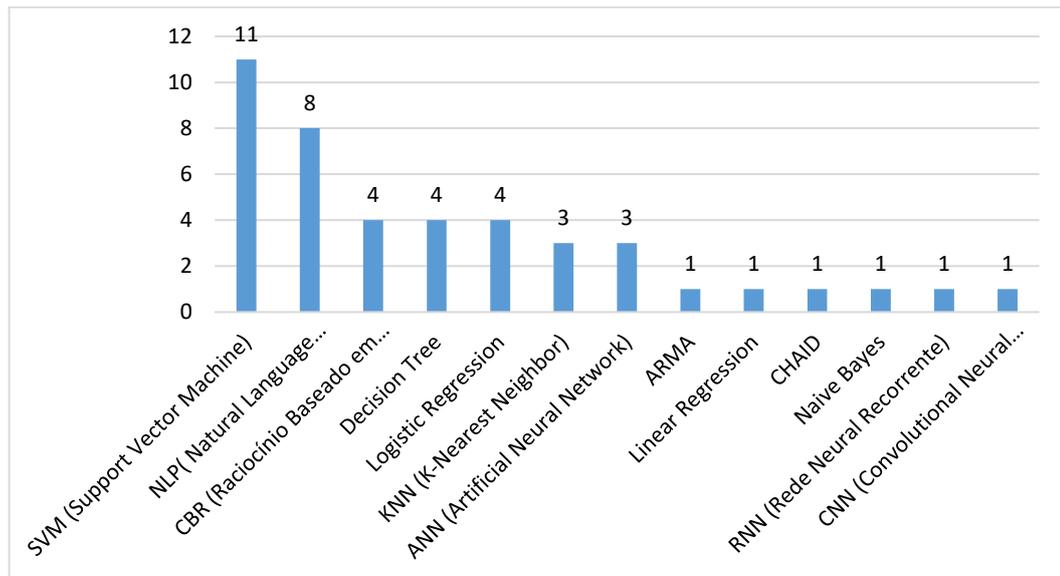
Técnica	Quantidade
SVM (Support Vector Machine)	11
NLP( Natural Language Processing)	8
CBR (Raciocínio Baseado em Casos)	4
Decision Tree	4
Logistic Regression	4
KNN (K-Nearest Neighbor)	3
ANN (Artificial Neural Network)	3
ARMA	1
Linear Regression	1
CHAID	1
Naive Bayes	1
RNN (Rede Neural Recorrente)	1
CNN (Convolutional Neural Network)	1

Fonte: Autor (2024)

Conforme Figura 8 as técnicas mais apontadas nos artigos pesquisados foram a máquina de suporte de vetores ou *Support Vector Machine* - SVM e o Processamento de Linguagem Natural ou natural language processing – NLP consideradas respectivamente em 11 e 8 artigos.

As técnicas citadas em mais de um artigo foram árvores de decisão (*Decision Tree*), Regressão Logística (*Logistic Regression*), K vizinhos mais próximos (*k-Nearest Neighbor*) e Redes Neurais Artificiais (*Artificial Neural Network*). Outras técnicas citadas em um único artigo foram ARIMA, Regressão Linear, CHAID, Naive Bayes, Redes Neurais Recorrentes e Redes Neurais Convolucionais”.

Figura 8 – Técnica utilizada versus quantidade



Fonte: Autor (2024)

Um exemplo de uso de técnicas em conjunto seria o trabalho representado pelo artigo "A Hybrid Continual Learning Approach For Efficient Hierarchical Classification Of It Support Tickets In The Presence Of Class Overlap" no qual é proposto um método híbrido baseado em classificador e um algoritmo baseado em regras. Primeiro é detectado as classes envolvidas usando a matriz de confusão do classificador. Em seguida são geradas N regras com intervenção mínima dos agentes de suporte. Finalmente, os tickets são enviados ao algoritmo para filtrar a classe a ser classificada com o classificador SVM. No que diz respeito à utilização de diversas técnicas pode-se facilmente concluir que o processamento de linguagem natural tem desempenhado um papel transformador no setor de suporte ao usuário de informática. A capacidade de compreender as consultas em linguagem natural, realizando tarefas como remoção de sufixos, eliminação de palavras de ruído e adição de sinônimos, representa um avanço significativo na melhoria do atendimento ao cliente e na eficiência do gerenciamento de incidentes.

A automação do helpdesk, oferecendo serviços na World Wide Web, tem se destacado como uma solução inovadora que impulsiona a economia e a interatividade no atendimento ao cliente haja vista muitas centrais de suporte atenderem clientes ao redor do mundo.

Nota-se também, a capacidade de prever respostas do *helpdesk* a partir de características presentes nos e-mails dos clientes e o prognóstico de solicitações de serviço em centros de suporte denotam avanços importantes que melhoram ainda mais a qualidade do suporte ao cliente e reduzem custos.

A classificação de problemas, e, em especial a facilitação do aconselhamento para analistas menos experientes tal como a análise de emoções, permitem um atendimento mais personalizado e eficaz. Não menos importante, temos também, o desenvolvimento de modelos de custo e a previsão de demanda como elementos cruciais para a alocação adequada de recursos.

A automação, a categorização e a previsão desempenham um papel fundamental na otimização do uso de recursos humanos da equipe de suporte de TI e na aceleração do atendimento ao cliente. A atribuição e encaminhamento de tíquetes, juntamente com a predição de escalonamento, resultam em resoluções mais rápidas e eficientes.

Em suma, o uso de diversas técnicas de ML têm revolucionado o setor de suporte ao usuário, oferecendo soluções abrangentes e eficazes que buscam a eficiência, reduzem custos e melhoram a satisfação do cliente. Essas inovações estão moldando o futuro do atendimento ao cliente, proporcionando um serviço mais personalizado e interativo.

Como o desenvolvimento do processo proposto nesse trabalho é baseado no método *DSR*, foi importante notar neste capítulo não só os conceitos e metodologia pertinentes, mas também, as três etapas iniciais do método. Nos resultados, seguindo a proposta do *DSR* teremos na discussão as próximas etapas.

#### **4. PROCESSO PROPOSTO (PRODIMAH)**

Considerando as etapas do modelo DSR adotado neste trabalho o presente capítulo traz a descrição do processo proposto e logo após teremos os resultados da pesquisa. Vale ressaltar, que as etapas iniciais de identificação e conscientização do problema, foram previamente tratadas na introdução e a revisão bibliográfica, devido à sua importância, foi apresentada no quarto capítulo.

##### **4.1 Identificação dos estudos e configuração das classes de problemas**

Ao identificar os estudos de acordo com Dresch *et al.*, (2015) é possível que o pesquisador se depare com um artefato ideal e pronto, e, sendo o caso, o pesquisador poderá continuar com a pesquisa à medida em que o novo artefato traga melhores soluções em comparação ao existente.

Neste estudo, considerando que as strings utilizadas trouxeram 1047 trabalhos na *Web of Science* e 1893 trabalhos da *Scopus* e em uma leitura inicial, constatou-se que muitos trabalhos não teriam relação com o objetivo da revisão sistemática da literatura. Foram aplicados os filtros conforme Figura 7 e não foi possível identificar nenhuma ferramenta ou estudo que possibilitasse atingir o objetivo deste estudo.

##### **4.2 Proposição do artefato para resolver o problema**

No contexto de manter o rigor metodológico e etapas da DSR Dresch *et al.*, (2015) apontam que o pesquisador poderá iniciar a etapa de proposição do artefato para resolução do problema após identificados os estudos que possivelmente configurem sua solução. Afirmam que a etapa de proposição de artefato é necessária, pois a identificação de classes de problemas e de artefatos desenvolvidos trata da visualização de possíveis artefatos. No entanto, tais soluções quando consolidadas precisam ser adaptadas a realidade do estudo propondo uma solução para o problema em questão.

Considerando a impossibilidade de encontrar um estudo que trouxesse uma proposta para elaborar uma ferramenta de auxílio na definição do SLA, o presente trabalho apresenta o

PRODIMAH (Processo para Diagnóstico e melhoria de atendimento em *helpdesk*), uma proposta de etapas para a criação de um fluxo de atividades que visam estabelecer um processo que permitirá o diagnóstico de atendimento de *helpdesk* e melhorar a definição de SLA se os dados apresentados pela empresa permitissem o uso de ML.

### 4.3 Projeto e desenvolvimento do artefato selecionado

Continuando as diretrizes propostas por Dresch *et al.*, (2015) para o projeto de artefatos e em seguida o seu desenvolvimento, tem-se que considerar as características internas do PRODIMAH, processo para diagnóstico e melhoria de atendimento em *helpdesk* e o contexto que ele irá operar, incluindo os componentes e suas relações de funcionamento.

Após concluída a fase do projeto temos de acordo com Dresch *et al.*, (2015) a sugestão para prosseguir com a fase de desenvolvimento, podendo nesta etapa, ser utilizadas diferentes abordagens como, por exemplo, algoritmos computacionais, representações gráficas, protótipos etc.

No presente estudo temos conforme Figura 9 a proposta de atividades necessárias para a construção do processo para diagnóstico e melhoria de atendimento em *helpdesk*.

A primeira atividade consiste na reunião com a empresa e tem como objetivo compreender o contexto da empresa, incluindo sua natureza de negócio, objetivos estratégicos, estrutura organizacional e cultura empresarial. As atividades envolvem identificar as expectativas, requisitos e prioridades da empresa, além de alinhar as entregas de TI com as metas organizacionais. Na sequência, a obtenção da base de dados tem como objetivo recolher dados dos atendimentos para análise posterior. A principal atividade é exportar um arquivo .csv do sistema atual que armazena os dados de atendimento possibilitando sua análise pela linguagem Python. Na sequência a atividade consiste na análise dos dados existentes utilizando estatística descritiva e tem como objetivo compreender a situação atual da prestação de serviços de *helpdesk* possibilitando criar relatórios e identificar tendências e áreas de melhoria. A revisão de SLA e métricas de desempenho possibilita adequar rapidamente às novas necessidades e ajustar as métricas de desempenho. Com a adequação e ajustes é possível a revisão dos processos e procedimentos objetivando melhorar a eficiência e qualidade dos serviços prestados. Na sequência a limpeza dos dados garante que os dados estarão prontos para análise

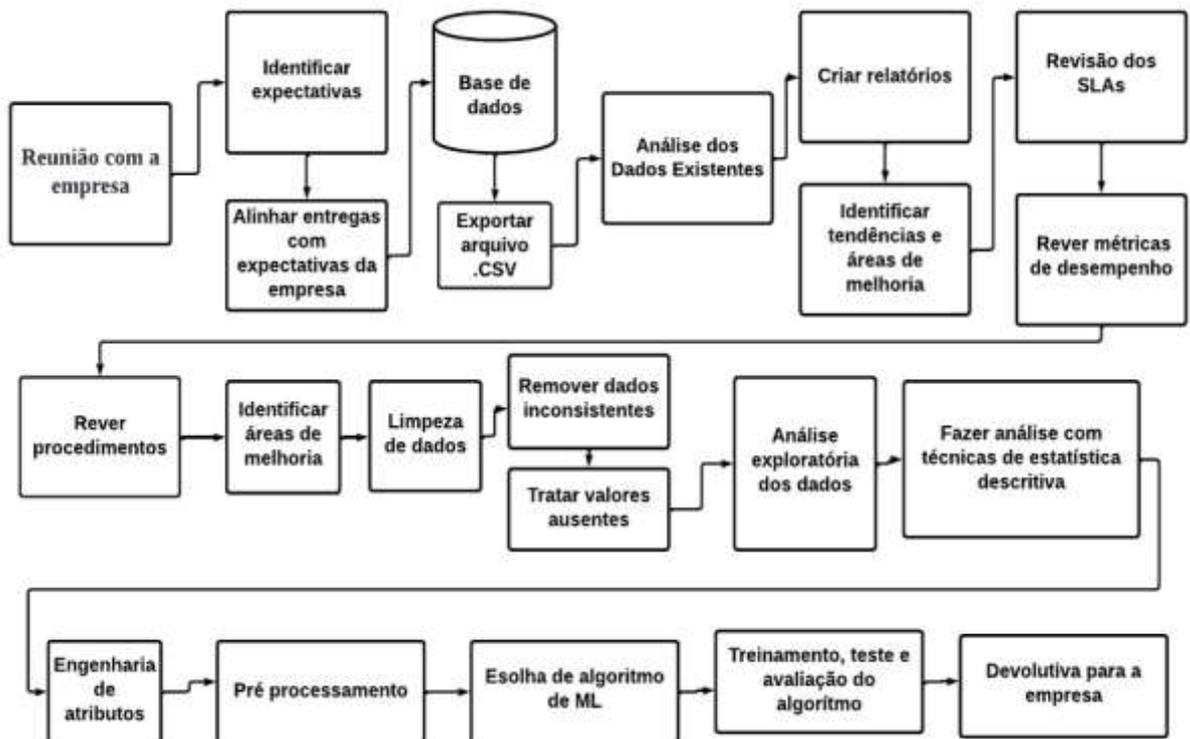
exploratória. As atividades envolvem remover dados inconsistentes ou inválidos e tratar valores ausentes e duplicados.

Com o objetivo de compreender a distribuição dos dados de atendimentos e identificar padrões temos a análise exploratória dos dados e suas atividades incluem utilizar técnicas de visualização de dados e descrever estatisticamente os dados. A engenharia de atributos possibilita a criação de novos atributos a partir dos dados existentes para melhorar a análise. Suas atividades envolvem criar variáveis derivadas e selecionar as características mais relevantes.

No pré-processamento dos dados o objetivo é preparar os dados para aplicação de modelos de *Machine Learning*. A atividade inclui divisão dos dados em conjuntos de treinamento e teste.

Aplicação de *Machine Learning* com o objetivo de prever a conclusão dentro ou fora do SLA e o tempo de atendimento. As atividades envolvem treinar modelos de ML, avaliar a performance dos modelos e implementar o modelo que apresentar melhor desempenho.

Figura 9 - Proposta para resolução



Fonte: Autor (2024)

Neste capítulo o projeto e desenvolvimento do artefato que consiste em um processo foi apresentado. A aplicação do processo e a avaliação feita sobre esta aplicação são apresentados no próximo capítulo denominado Resultados.

## 5. RESULTADOS

Considerando as etapas do modelo DSR adotado neste trabalho o presente capítulo traz os resultados da pesquisa. Vale ressaltar, que as etapas iniciais de identificação e conscientização do problema foram previamente tratadas na introdução, a revisão bibliográfica, devido à sua importância, foi apresentada no terceiro capítulo e a proposição, projeto e desenvolvimento do processo proposto foram apresentados no capítulo 4. Como resultados, neste capítulo, serão apresentadas a aplicação do processo proposto e a avaliação do processo a partir desta aplicação.

### 5.1 Aplicação do processo proposto

Conforme havia sido previsto na metodologia deste trabalho após a proposição, projeto e desenvolvimento do artefato que neste caso é o processo como solução para diagnóstico e melhoria do atendimento de *helpdesk* que prevê o uso de estatística descritiva e ML, o processo foi aplicado para a validação em uma empresa de capital 100% nacional, especializada em TI e integradora de soluções cujo principal foco é o fornecimento de soluções integradas em serviços de TI, *softwares*, aplicativos e sistemas de gestão empresarial, empregando as mais conceituadas plataformas globais para os setores de Agronegócio, Manufatura, Automotivo, Logística, Serviços e Varejo.

O contato com a empresa se deu por meio virtual onde foram apresentados os responsáveis pela operação de atendimento e nesta reunião foi designado como contato para disponibilização da base de dados o administrador do banco de dados e no momento foi apontada a necessidade de comunicação mesmo durante as etapas posteriores em função da necessidade de ajuste das expectativas quanto ao nível de atendimento pelo *helpdesk*. Na reunião foi importante compreender o contexto da empresa, incluindo sua natureza de negócio, objetivos estratégicos, estrutura organizacional e cultura empresarial. Foi possível identificar as expectativas, requisitos e prioridades da empresa em relação aos atendimentos, garantindo assim o alinhamento entre as entregas de TI e as metas organizacionais.

A atividade para a obtenção da base de dados sobre os atendimentos, se deu pela exportação de um arquivo com extensão “.csv” do sistema atual que armazena os dados.

Foi informado que a análise dos dados existentes utilizaria estatística descritiva e que serviria para compreender a situação atual da prestação de serviços de *helpdesk*, possibilitando a confecção de relatórios que viabilizariam a identificação de tendências, por exemplo.

Outro ponto importante a ser considerado foi a possibilidade de revisar o SLA em virtude da necessidade de rápida adaptação às novas necessidades existentes assim como adequar as métricas de desempenho pré-estabelecidas para medir a qualidade e eficiência dos serviços prestados. Além disso, haveria a possibilidade de revisar os processos e procedimentos existentes para a entrega e suporte de serviços de TI, identificando áreas que possam ser otimizadas ou aprimoradas para melhorar a eficiência e qualidade dos serviços.

A etapa técnica, importante em especial para a verificação da possibilidade do uso de ML, aconteceu entremeada à etapa administrativa à medida que esta etapa necessita de análise de dados para a compreensão da situação atual, diagnóstico e estudo de viabilidade de aplicação de ML para predição de conclusão e tempo de atendimento. Nessa macro etapa considera-se a limpeza dos dados, análise exploratória dos dados, engenharia de atributos, pré-processamento eventual aplicação de ML.

A conclusão da a etapa para obtenção da base de dados possibilitou analisar as informações efetivamente coletadas sobre os serviços prestados, incluindo dados de atendimento, métricas de desempenho, tendências e áreas de melhoria nos serviços de TI. Vale observar que não foi possível analisar o feedback dos usuários.

Considerando que o arquivo fornecido é de extensão “.csv” foi definida a ferramenta para análise dos dados e percebe-se na Figura 10 a versão 3.9.7 utilizada do software de desenvolvimento Python.

Figura 10 - Versão utilizada do Python

```
# Versão da Linguagem Python
from platform import python_version
print('Versão da Linguagem Python Usada Neste Jupyter Notebook:', python_version())
```

Versão da Linguagem Python Usada Neste Jupyter Notebook: 3.9.7

Fonte: Autor (2024)

O enfoque detalhado na seleção de ferramentas e tecnologias facilita a compreensão do contexto em que a ferramenta opera, proporcionando uma base sólida para a avaliação do desempenho e a replicabilidade.

Figura 11 - Bibliotecas utilizadas

```
In [2]: # Imports
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")

In [3]: # Versões dos pacotes usados neste jupyter notebook
%reload_ext watermark
%watermark -a "William Egidio de Noronha" --iversions

Author: William Egidio de Noronha

matplotlib: 3.4.3
seaborn : 0.11.2
pandas : 1.4.2
numpy : 1.22.3
```

Fonte: Autor (2024)

Também foram definidas as bibliotecas Numpy, Pandas, Matplotlib, Pyplot e Seaborn, conforme apresenta a

Figura 11, a fim de permitir o tratamento eficaz de dados e a manipulação matemática necessária para a operacionalização e apresentação dos gráficos.

Na sequência, após a importação da base de dados foram conhecidas as colunas do *dataset*, sendo elas representadas pela Figura 12.

Figura 12 - Colunas do *dataset*

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1918 entries, 0 to 1917
Data columns (total 21 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Identificação da solicitação             1884 non-null   float64
1   Departamento                             1884 non-null   object
2   Categoria                                 1884 non-null   object
3   Subcategoria                             1884 non-null   object
4   Assunto                                   1884 non-null   object
5   Criado por                               1884 non-null   object
6   Técnico                                  1884 non-null   object
7   Descrição                                 1884 non-null   object
8   Local                                     1884 non-null   object
9   Status da solicitação                   1884 non-null   object
10  Status pendente                          1884 non-null   object
11  Hora de criação                          1884 non-null   object
12  Data de resposta                          1884 non-null   object
13  Hora do vencimento                       1884 non-null   object
14  Hora de conclusão                        1884 non-null   object
15  Tempo decorrido                          1884 non-null   object
16  Status do vencimento                     1884 non-null   object
17  Resposta Devida às                       1884 non-null   object
18  Status da Primeira Resposta Atrasada    1884 non-null   object
19  Horário da última atualização            1884 non-null   object
20  Hora da solução                          1884 non-null   object
dtypes: float64(1), object(20)
memory usage: 314.8+ KB
```

Fonte: Autor (2024)

Foram identificadas 21 colunas com 1884 linhas. A definição de cada coluna segue conforme abaixo:

- Departamento: Representa o departamento em que o atendimento foi provido;
- Categoria: Determina a categoria macro onde a solicitação será classificada;
- Subcategoria: Determina dentro da categoria a subcategoria respectiva a solicitação;
- Criado por: Determina o nome do analista responsável pela abertura da solicitação no momento do acionamento;
- Técnico: Determina o nome do técnico responsável pelo atendimento a solicitação;
- Local: Determina a localização da ocorrência;
- Status da Solicitação: Determinar a situação da solicitação em um momento específico, podendo ser Aberto, Em atendimento técnico, Encaminhado ao Solicitante, Em homologação Aguardando aprovação e Resolvido.
- Tempo decorrido: Indica o tempo decorrido até o momento da verificação;
- Status do vencimento: Mostra se a solicitação foi atendida dentro ou fora do prazo estipulado pelo SLA.
- Status da primeira resposta atrasada: Com valor verdadeiro ou falso a coluna denota, quando o status é verdadeiro, que não houve contato inicial no prazo de 4 horas.
- Identificação da solicitação: Número que identifica o atendimento como único;
- Assunto: Descreve o assunto de que se trata a solicitação em texto livre.
- Descrição: Coluna para inserção de texto corrido contendo a descrição do atendimento;
- Status pendente: Determina se o atendimento no momento depende de algum fornecedor ou resposta por parte do solicitante;
- Hora de criação: Determina a hora de criação do chamado;
- Data de resposta: Determina a data em que primeira resposta do analista responsável foi dada;
- Hora do vencimento: Momento em que o chamado terá seu SLA vencido;
- Hora da conclusão: Horário em que o chamado foi de fato concluído;

- Status do vencimento: Determina se o chamado foi atendido ou não dentro do prazo.
- Hora da última atualização: Horário da última atualização do chamado feita pelo analista responsável;
- Hora da solução: Horário da solução apresentada pelo analista.

Após identificadas foram definidas as colunas Departamento, Categoria, Subcategoria, Criado por, técnico, Local, status da solicitação, Status do vencimento, Status da primeira resposta atrasada, Identificação da solicitação, Criado por e Status do vencimento como principais haja vista representarem variáveis categóricas que possibilitariam o uso em algoritmos de ML. Concluída esta etapa foi definido o fluxo de atividades composto pela análise inicial dos dados, processo de limpeza, análise exploratória dos dados, engenharia de atributos e por fim o pré-processamento.

### 5.1.1 Análise inicial dos dados

As técnicas que nos proporcionam analisar e interpretar as informações básicas dos dados coletados de acordo com Pereira *et al.*, (2019) fazem parte de um conjunto chamado de Investigação ou Estatística Descritiva. A descrição dos dados é fornecida pelas medidas de posição, também chamadas de tendência central, as quais apresentam a frequência dos dados que nos dizem o quão dispersos ou distantes um do outro estão os valores de um conjunto de dados e são chamadas de variância e desvio padrão.

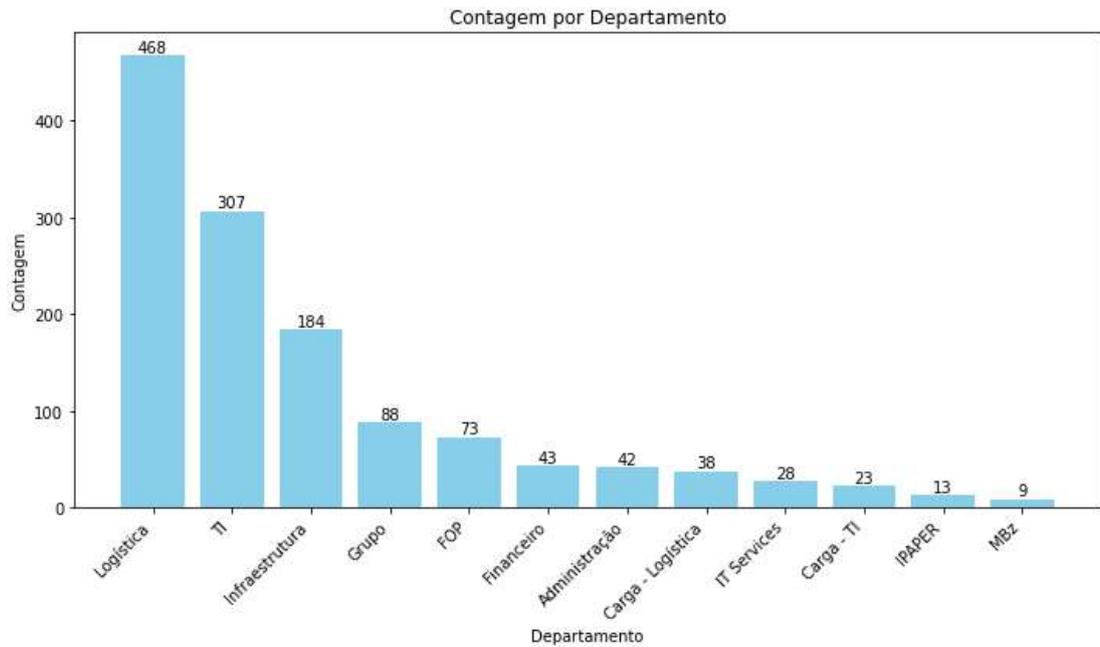
Na análise inicial dos dados foi percebida a média de tempo de atendimento de 34,50 horas com desvio padrão de 135,30 horas. Também foi verificado que 25% do total de linhas estavam preenchidas com 0 horas de atendimento. Entre 0 e 50% do total de linhas estavam com preenchimento de até 2 horas. Até 75% foi percebido o tempo de 15 horas e o tempo máximo de atendimento foi de 2351 horas.

Na coluna departamento, representada pela

Figura 13 **Erro! Fonte de referência não encontrada.**, pode ser observado que as linhas TI e IT Services e Tecnologia da Informação sugerem ser do mesmo departamento assim como “Administração” e “Administracion”. Logística, TI, Infraestrutura e não atribuído, têm seus

destaques pelo volume total e representação numérica em relação ao Departamento com menor número de chamados.

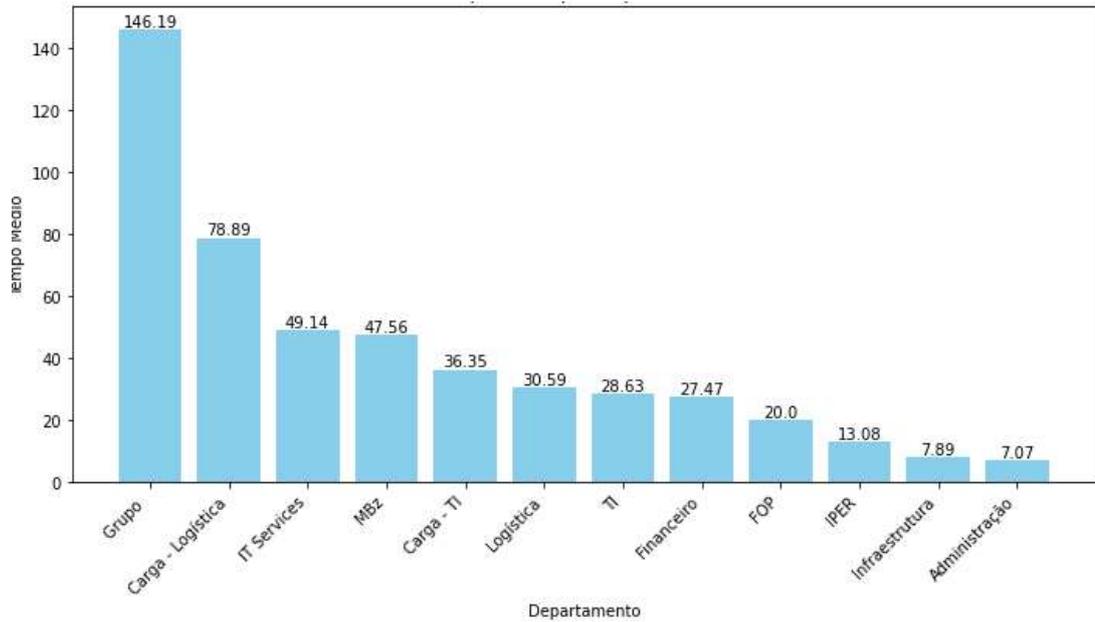
Figura 13 - Análise inicial da coluna Departamento



Fonte: Autor (2024)

Na coluna Departamento representada pela Figura 14 foi possível identificar as duas maiores médias de tempo de atendimento com Grupo e Carga Logística que não estão entre os departamentos que atenderam as maiores quantidades de solicitações.

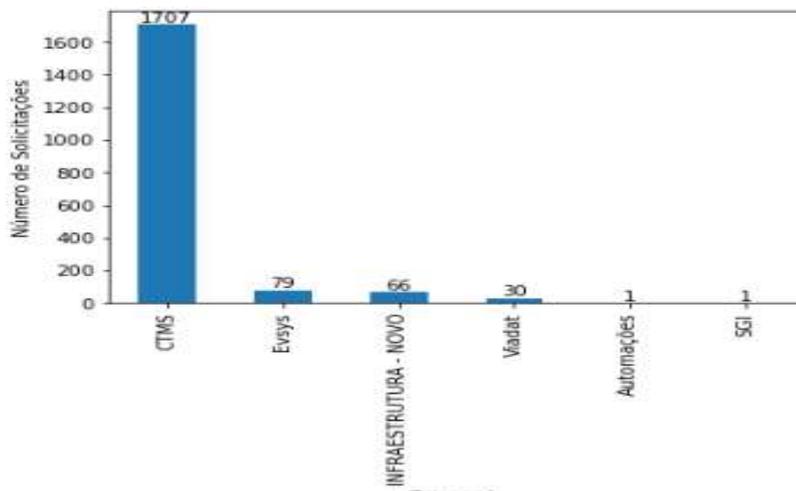
Figura 14 - Tempo médio por Departamento



Fonte: Autor (2024)

Na coluna Categoria, representada pela Figura 15, pôde-se observar a existência de seis categorias com Automações e SGI contendo valores minimamente significativos. CTMS demonstrando ser a categoria mais representativa em volume de chamados com diferença significativa em relação às categorias menos representativas numericamente.

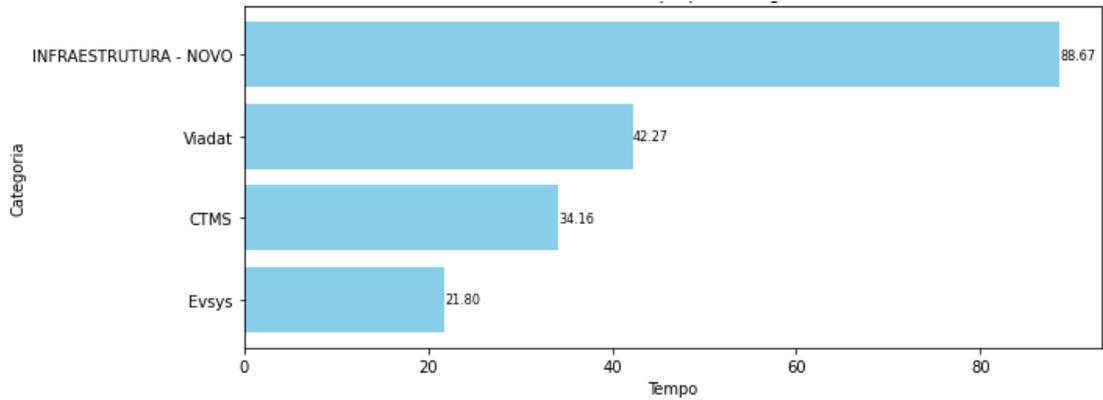
Figura 15 - Análise inicial da coluna Categoria



Fonte: Autor (2024)

No que diz respeito à média do tempo de atendimento é possível identificar pela Figura 16 que INFRAESTRUTURA-NOVO seguida de Viadat ainda que com pouco volume nos solicitações apresentam as maiores médias de tempo de atendimento com 88,67 e 42,27 horas respectivamente.

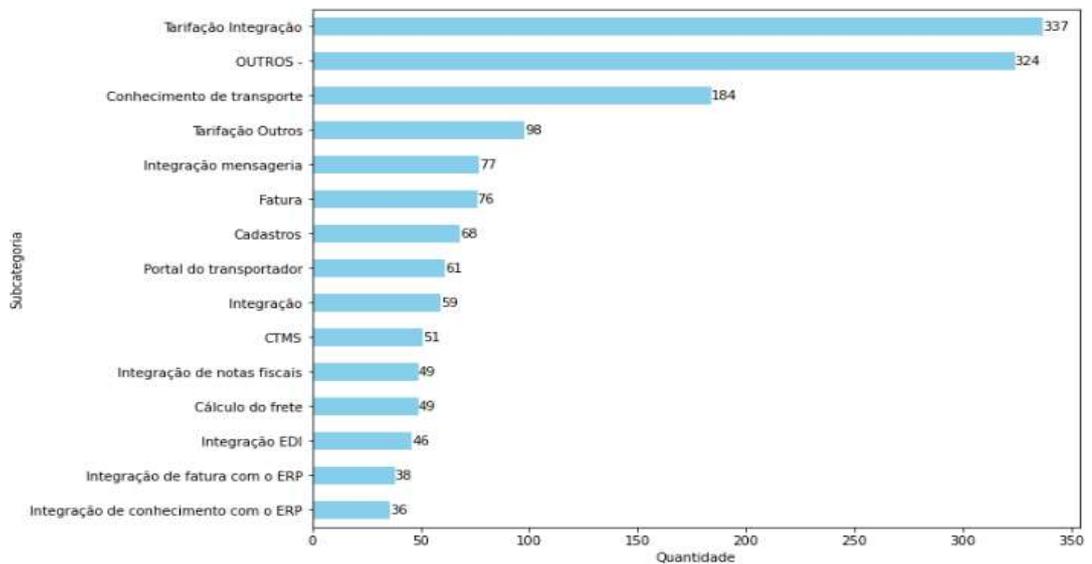
Figura 16 - Tempo médio por Departamento



Fonte: Autor (2024)

Prosseguindo pela análise inicial e considerando a coluna Subcategoria, foram encontradas 44 categorias e pode ser observado pela Figura 17 a subcategoria Outros com representativo volume na quantidade de solicitações. Assim como nas colunas anteriores, também é verificada diferença significativa entre a subcategoria de maior volume, como Tecnologia da informação, Outros e Conhecimento de transporte e menor volume, como por exemplo, Integração de conhecimento com ERP.

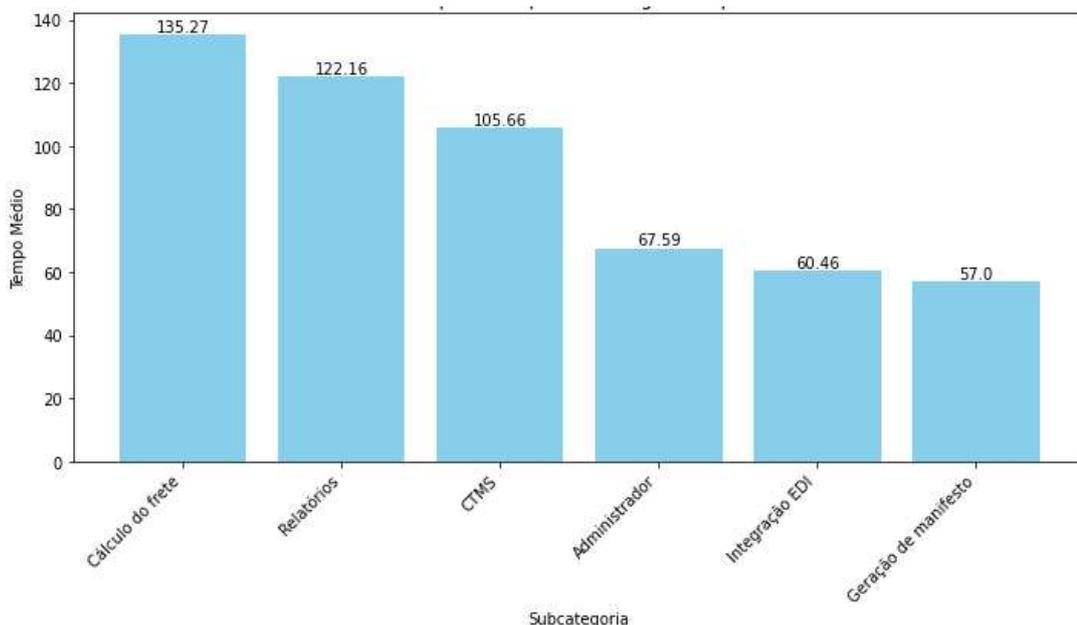
Figura 17 – Análise inicial Subcategoria



Fonte: Autor (2024)

Verificando as 6 maiores médias na Figura 18 é notado que cálculo do frete e relatórios têm destaque ainda que sejam subcategorias com poucas solicitações.

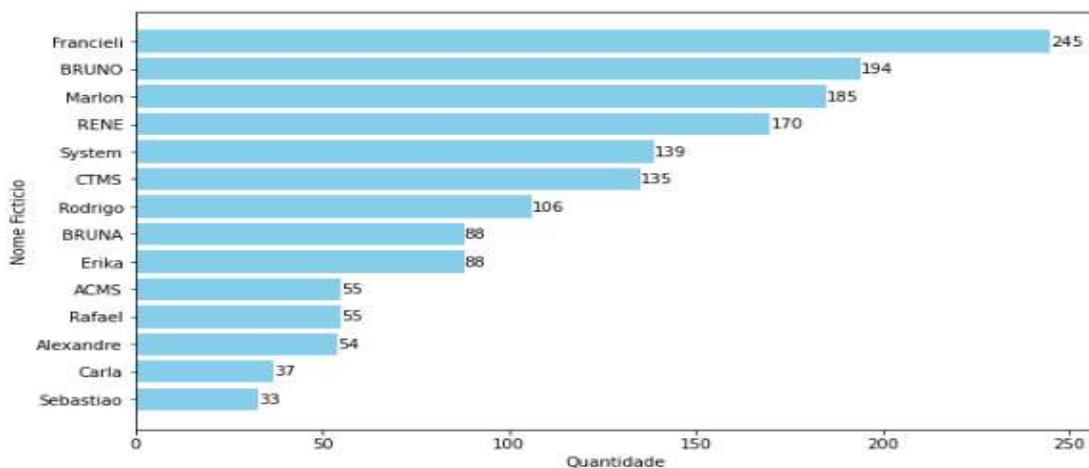
Figura 18 - Tempo médio por Subcategoria



Fonte: Autor (2024)

Na coluna Criado por, representada pela Figura 19 é possível identificar diferença significativa entre o analista que criou o maior número de solicitações e o analista que criou o menor número de solicitações, neste caso foi considerado apenas casos com quantidade superior a 30 registros no sistema haja vista ter sido percebida a característica de grande diferença entre os analistas que mais registraram e os que menos registraram, demonstrando valores de pouca representatividade, neste caso com um registro.

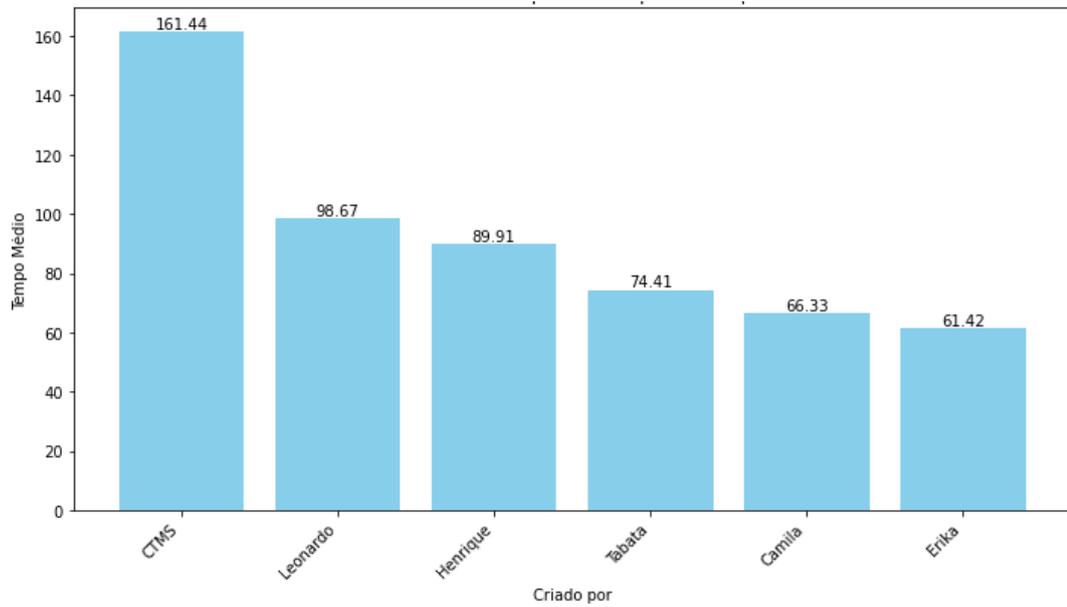
Figura 19 - Análise inicial Criado por



Fonte: Autor (2024)

No que diz respeito às seis maiores médias de tempo de atendimento de Criado por CTMS e Leonardo são destacados em especial por não terem grande volume de solicitações percebido na figura Figura 20.

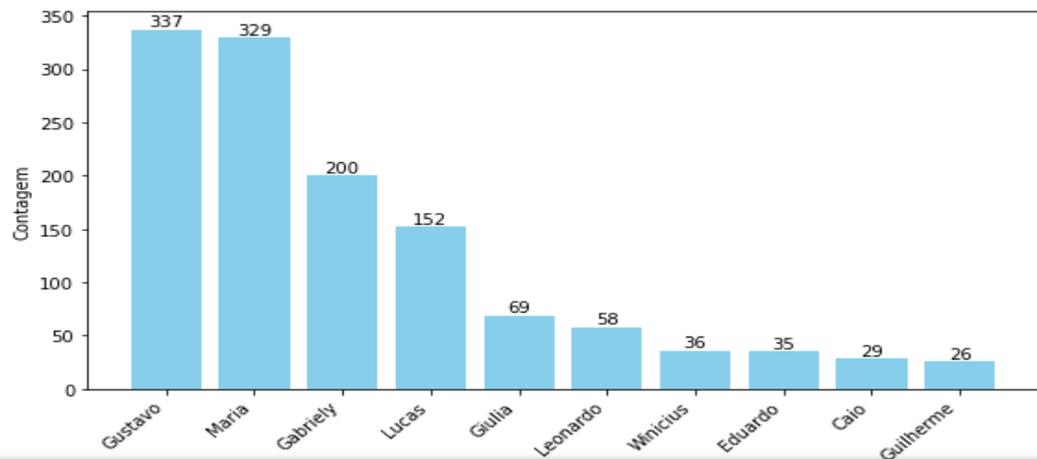
Figura 20 - Seis maiores tempo médio Criado por



Fonte: Autor (2024)

Característica semelhante também pode ser percebida na análise dos dados que representam os técnicos. É percebido na Figura 21 que o técnico que mais atendeu solicitações atingiu o número de 337 enquanto o técnico que menos atendeu solicitações está representado por apenas 26 solicitações, observa-se que não foram considerados os técnicos com menos de 20 solicitações.

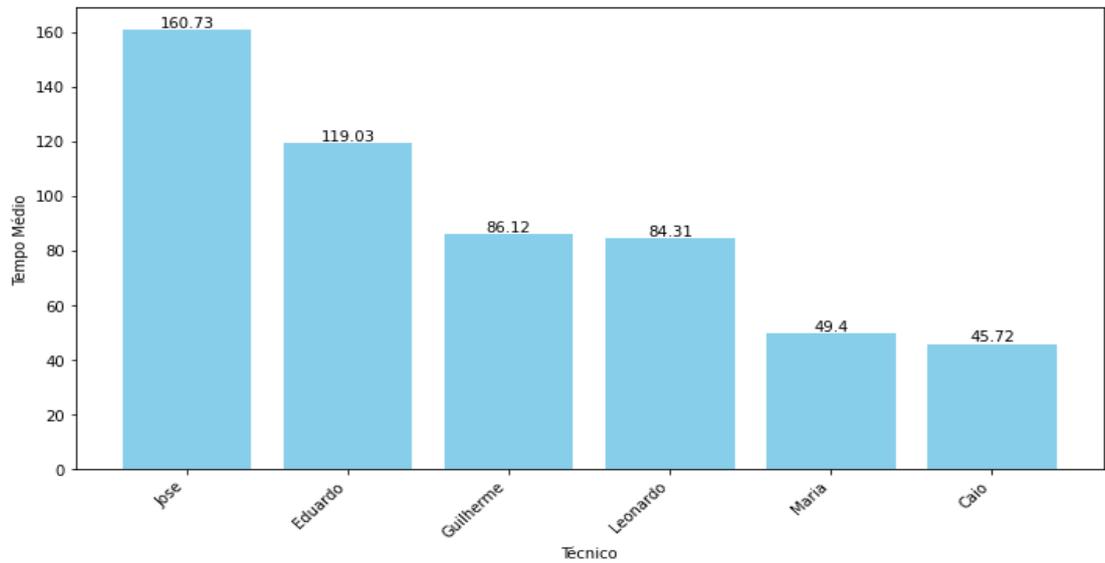
Figura 21 - Análise inicial Técnico



Fonte: Autor (2024)

Para as seis maiores médias de atendimento pode ser verificado na Figura 22 que Jose ainda que não tenha absorvido o volume de pelo menos 20 solicitações no período conseguiu seu destaque como maior média de tempo de atendimento.

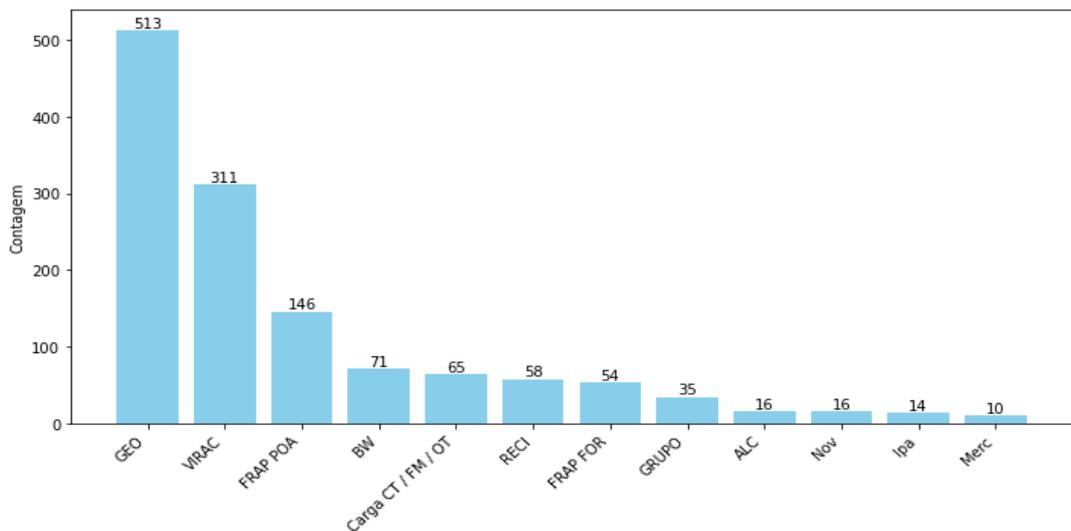
Figura 22 - Seis maiores médias de tempo por técnico



Fonte: Autor (2024)

A característica de ter muitas solicitações para algumas categorias e poucas para outras também foi percebida na Figura 23 representando as localidades.

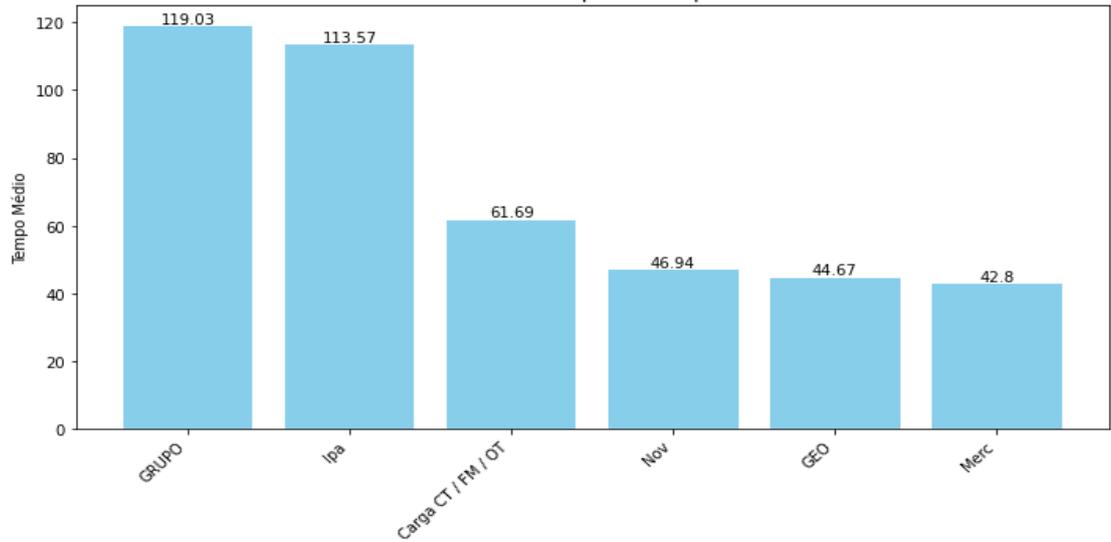
Figura 23 - Análise inicial da coluna Local



Fonte: Autor (2024)

Também foi apresentado os maiores tempos médios de atendimento para categorias de pouca representatividade na quantidade. Conforme Figura 24 pode ser observado que GRUPO E Ipa se destacaram.

Figura 24 - Seis maiores médias de tempo por local

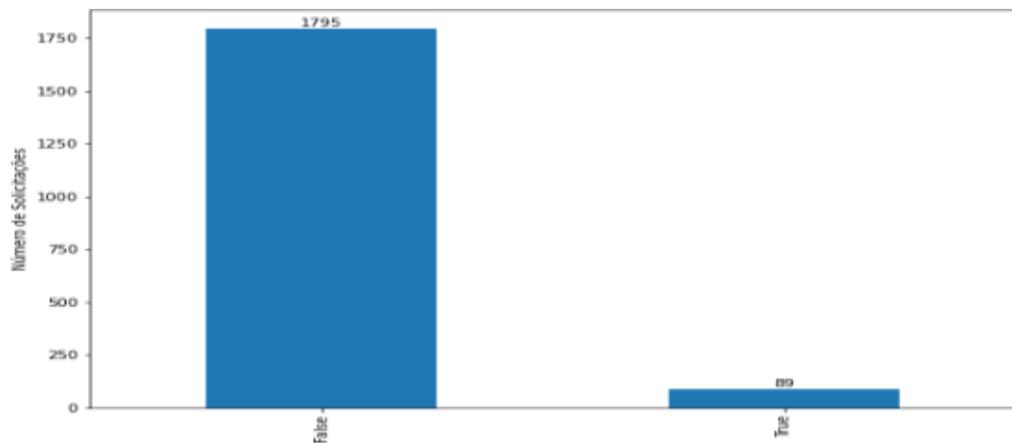


Fonte: Autor (2024)

Na

Figura 25 percebemos o número de 1795 solicitações que não tiveram a primeira resposta atrasada contrapondo 89 solicitações com a primeira resposta atrasada.

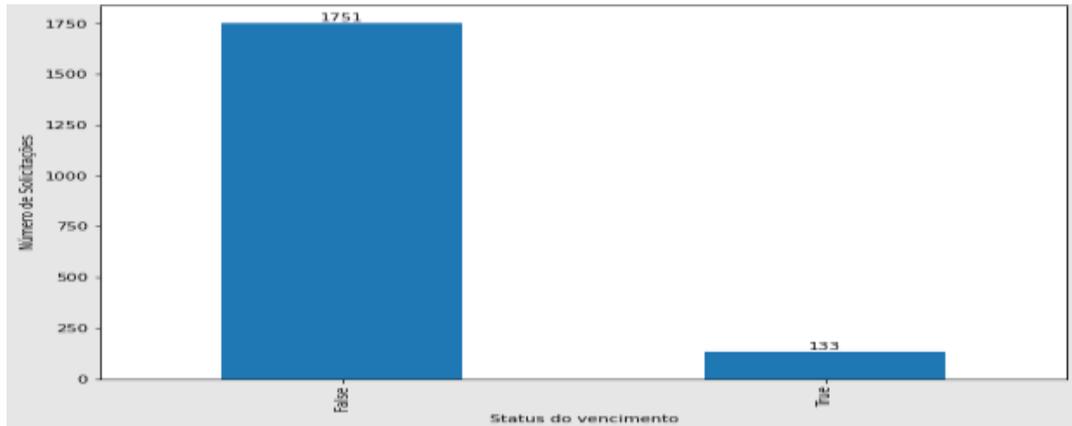
Figura 25 - Análise inicial da coluna primeira resposta atrasada



Fonte: Autor (2024)

A coluna Status do vencimento representada pela Figura 26 também merece destaque pois é objetivo do estudo prever o status do vencimento quando da finalização da categorização do chamado pelo analista responsável. Nesta coluna foi possível perceber 1751 solicitações que foram atendidas dentro do prazo e 133 solicitações que não foram atendidas dentro do prazo.

Figura 26 - Análise inicial da coluna Status do vencimento



Fonte: Autor (2024)

A análise inicial das colunas da base de dados possibilitou encontrar características relevantes e a conclusão da atividade possibilitou o início do processo para a limpeza dos dados.

### 5.1.2 Processo de limpeza dos dados

Foi possível constatar a presença da notação “NaN” (Not a Number) em diversas linhas e colunas do dataset, indicando a ausência de dado ou informação em determinados campos.

Observada a necessidade de limpeza de dados, foi necessário identificar a quantidade de linhas com informações ausentes. Para tanto, foi utilizado o comando `df.isna().sum()` para soma desses valores em cada coluna.

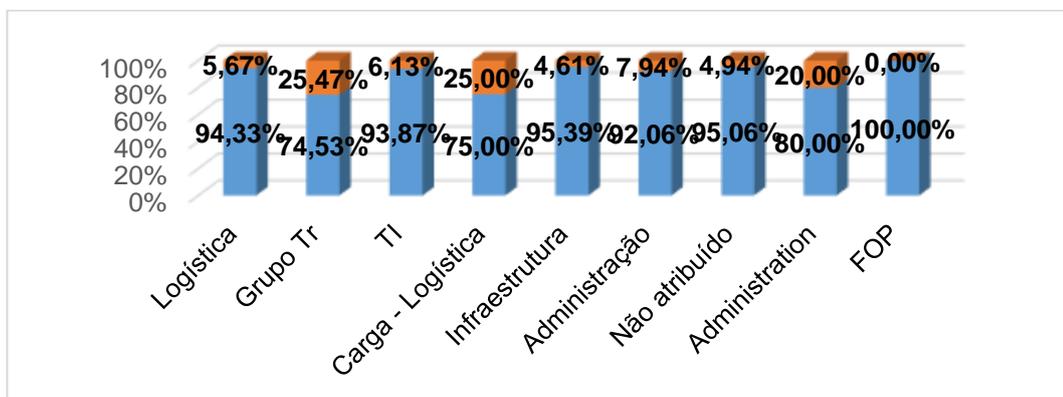
Foram identificadas 34 linhas em cada uma das colunas do data frame o comando `df.dropna()` para a exclusão foi utilizado e então foi verificado que a limpeza foi realizada.

Nessa fase também foi verificado que não havia linhas duplicadas. Seguindo o fluxo a próxima etapa discorre sobre a análise exploratória dos dados.

### 5.1.3 Análise exploratória dos dados

No mérito de entender os dados e com o intuito de verificar a quantidade de categorias com suas respectivas frequências no data frame, foi feita a contagem de valores em cada departamento contrapondo com a quantidade de solicitações atendidas e não atendidas. A notação false, cor laranja, para solicitações não atendidas dentro do prazo e a notação true, cor azul, foi utilizada para classificar as solicitações atendidas dentro do prazo, conforme Gráfico 1. Inicialmente foram consideradas as solicitações para a coluna dos departamentos com pelo menos 40 solicitações haja vista a presença de departamentos com pouca representatividade numérica.

Gráfico 1 - Coluna Departamento com percentual



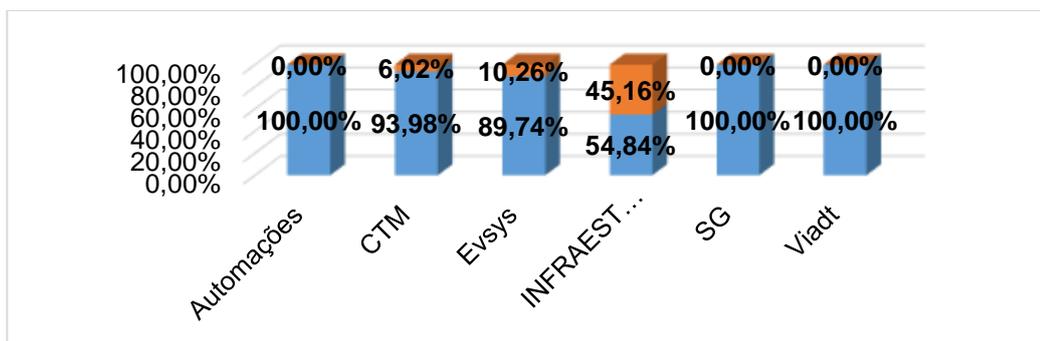
Fonte: Autor (2024)

Percebeu-se que o “Grupo Tr” e Departamento “Carga – Logística” são os mais representativos em termos percentuais de solicitações não atendidas dentro do prazo.

Vale observar que a notação *true* e *false* para solicitações atendidas ou não no prazo com suas respectivas cores foi replicada para as demais colunas.

Na análise da coluna Categoria representada pela Gráfico 2 o destaque foi observado pela categoria INFRA-ESTUTURA-NOVO com 45,16% das solicitações não atendidas dentro do prazo seguida da categoria Evsys.

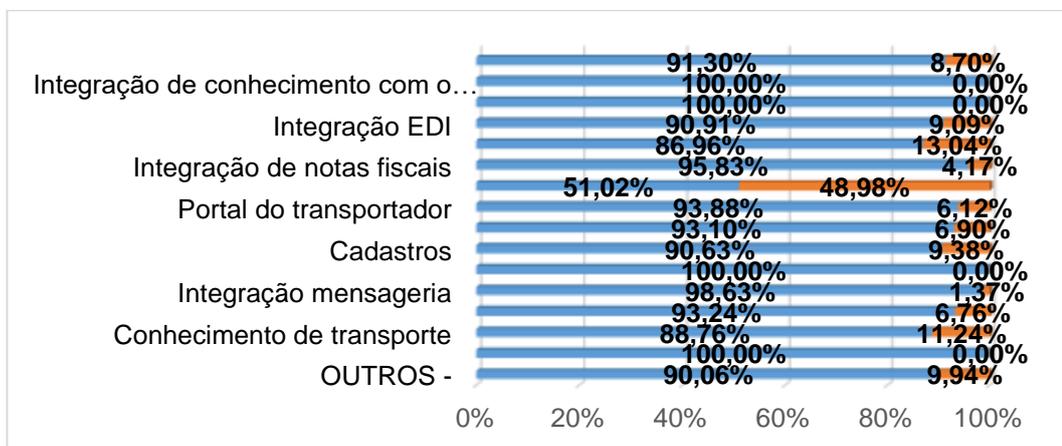
Gráfico 2 - Coluna Categoria com percentual



Fonte: Autor (2024)

Na sequência conforme Gráfico 3 para a coluna subcategoria foi considerado os atendimentos nas subcategorias com pelo menos 40 atendimentos. Em termos percentuais, foi percebido que a Subcategoria CTMS apresenta 48,98% das solicitações não atendidas dentro do prazo. Conhecimento de transporte é apresentada com 11,24% de solicitações não atendidas dentro do prazo. Outros e OUTROS - também aparecem com significativa representatividade entre as solicitações não atendidas dentro do prazo.

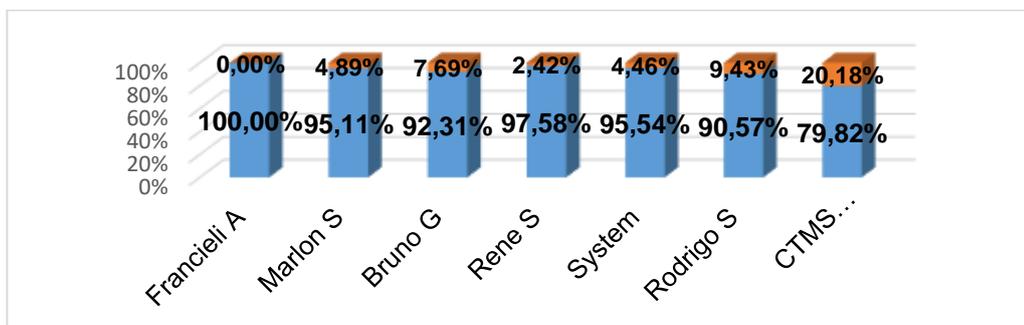
Gráfico 3 - Subcategoria por percentual



Fonte: Autor (2024)

Considerando a coluna “Criado por” percebeu-se, conforme Gráfico 4, por foram considerados os 7 maiores resultados em valor numérico. É apresentado como destaque de solicitações não atendidas dentro do prazo a categoria CTMS Atendimento com 20,18% de solicitações não atendidas dentro do prazo.

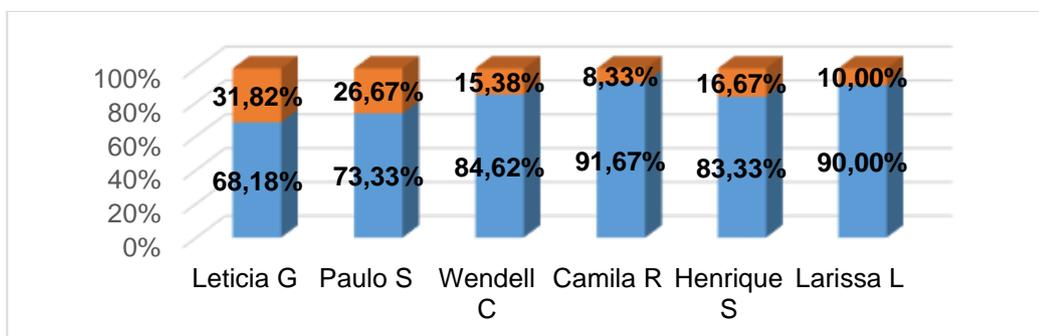
Gráfico 4 - Criado por com mais de 80 solicitações em percentual



Fonte: Autor (2024)

Considerando a coluna “Criado por” foram considerados os responsáveis com menor número representativo e ainda assim foi observado boa representatividade de solicitações que não foram atendidas dentro do prazo conforme Gráfico 5.

Gráfico 5 - Criado por até 40 solicitações

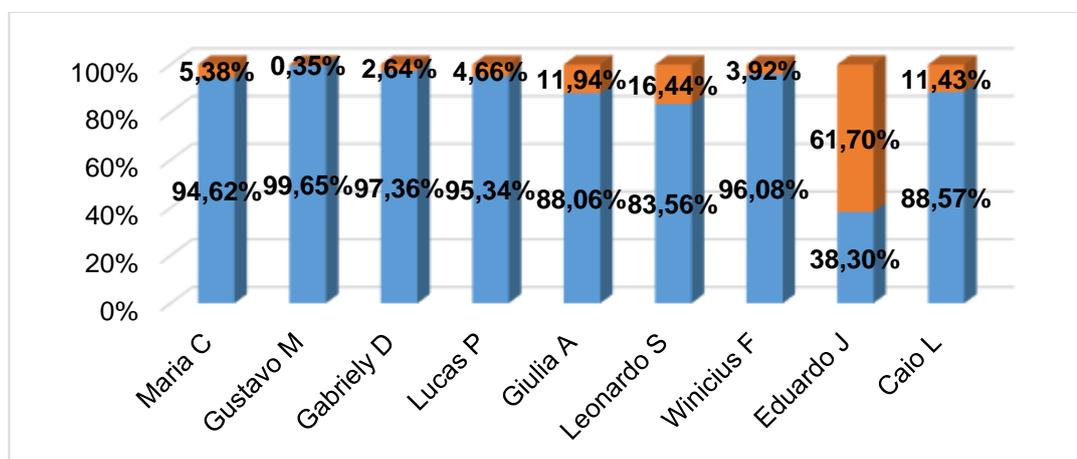


Fonte: Autor (2024)

Na análise da coluna Técnico, representada pela

Gráfico 6, foram selecionados os analistas que se destacaram pela quantidade percentual de solicitações não atendidas dentro do prazo conforme e atendimento para pelo menos 50 solicitações.

Gráfico 6 - Técnico com mais de 50 solicitações

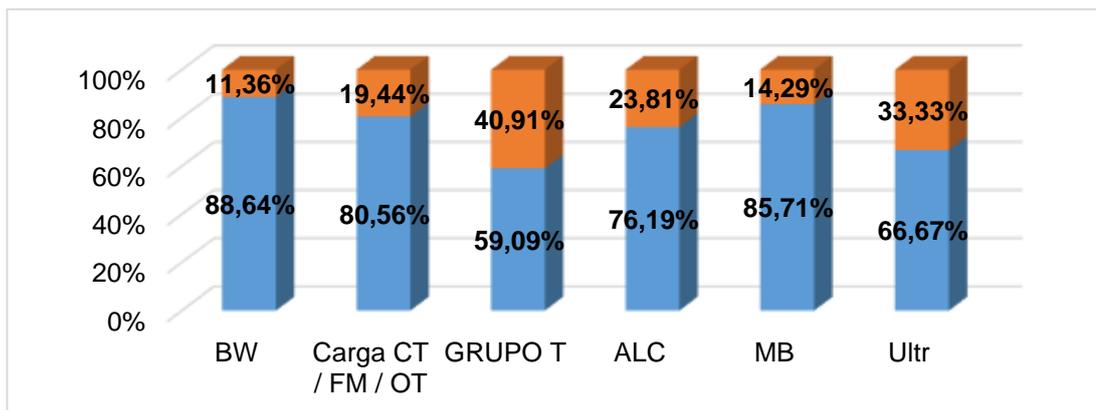


Fonte: Autor (2024)

Na coluna técnico vale destacar o técnico Eduardo J com 61,7% de suas solicitações não atendidas dentro do prazo.

No que tange à coluna Local o destaque fica para os locais demonstrados pelo Gráfico 7 que apresentaram a maior representatividade percentual de solicitações não atendidas dentro do prazo com destaque para o Grupo T.

Gráfico 7 - Local em percentual

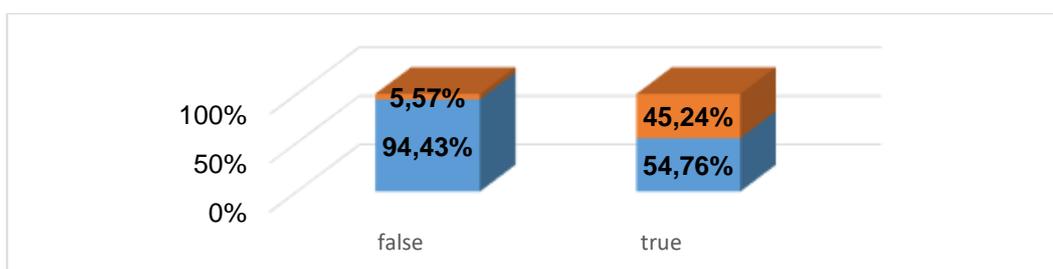


Fonte: Autor (2024)

Na análise da coluna Status da primeira resposta atrasada classificada como true e representada pelo

Gráfico 8 abaixo, é verificado o significativo percentual de solicitações não atendidas dentro do prazo.

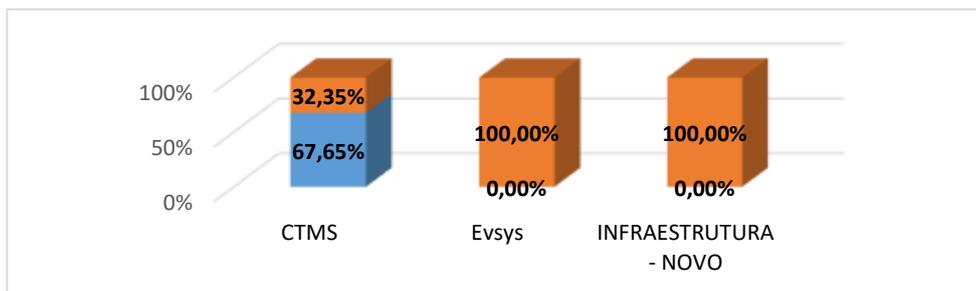
Gráfico 8 - Primeira resposta atrasada em percentual



Fonte: Autor (2024)

Com o intuito de verificar as colunas seleccionadas e que apresentaram tempo de atendimento superior a 50 horas vale observar a presença das Categorias CTMS, Evsys e INFRAESTRUTURA – NOVO, conforme Gráfico 9 abaixo.

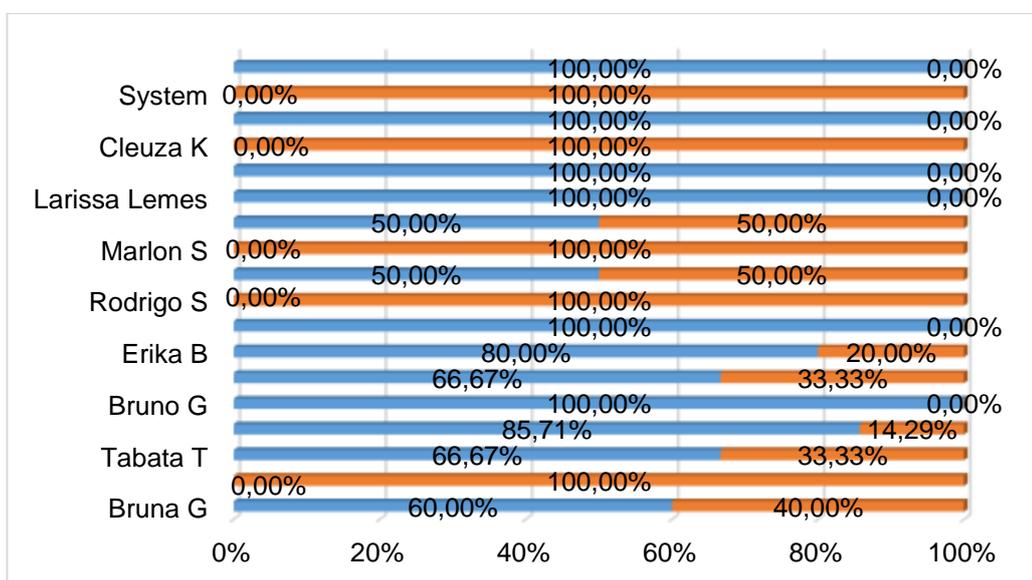
Gráfico 9 - Categoria com mais de 50 horas



Fonte: Autor (2024)

Permanecendo a perspectiva de solicitações com mais de 50 horas foi verificada a coluna “Criado por” conforme Gráfico 10 abaixo.

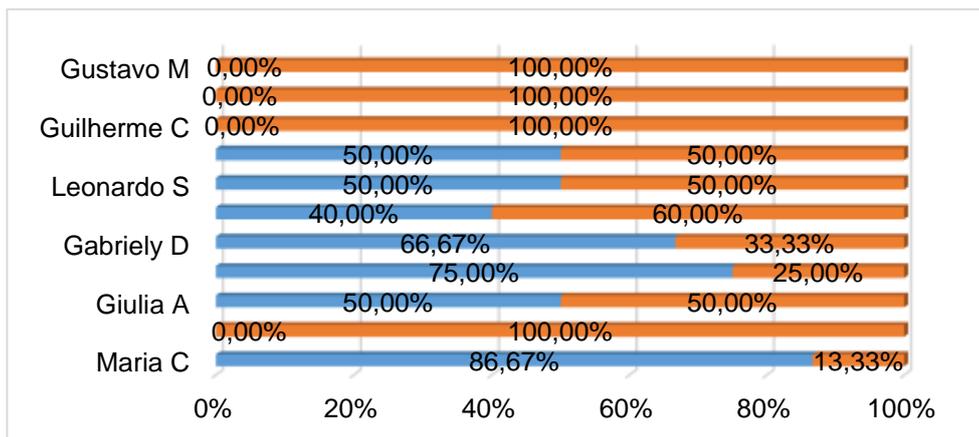
Gráfico 10 - Criado por com mais de 50 horas



Fonte: Autor (2024)

Considerando a coluna técnico com tempo de atendimento superior a 50 horas de acordo com o Gráfico 11 foi possível observar 9 técnicos com pelo menos 30% das solicitações não atendidas dentro do prazo.

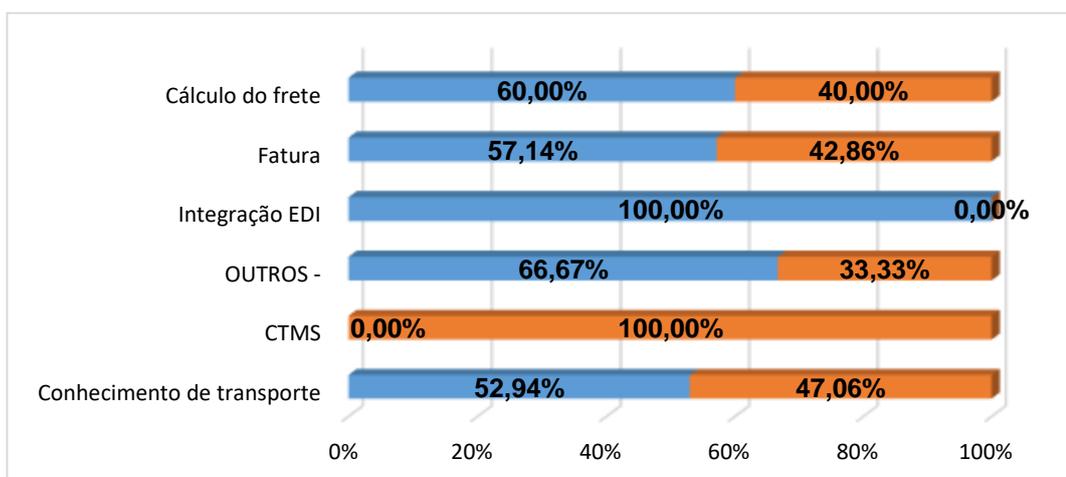
Gráfico 11 - Técnico com tempo de atendimento superior a 50 horas



Fonte: Autor (2024)

Na sequência e considerando as seis subcategorias que mais se destacaram em termos numéricos cujo tempo de atendimento foi superior a 50 horas temos conforme o Gráfico 12 abaixo que 5 subcategorias ainda que com o tempo superior a 50 horas também apresentaram percentual superior a 30% de solicitações não atendidas dentro do prazo.

Gráfico 12 - Seis primeiros da subcategoria com mais de 50 horas



Fonte: Autor (2024)

Na análise exploratória também foi verificada alta variabilidade de tempo. Ainda que tenham sido consideradas os mesmos departamentos, categorias, subcategoria, analista responsável pela criação da solicitação, técnico responsável pelo atendimento e local foi percebida significativa variação no tempo de atendimento, conforme Quadro 9.

Quadro 9 - Variação no tempo decorrido para atendimento

Departamento	Categoria	Subcategoria	Criado por	Técnico	Local	Tempo decorrido
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	135:33:05
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	113:09:50
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	100:31:46
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	66:48:52
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	34:26:55
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	23:28:08
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	23:25:03
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	20:34:47
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	19:00:24
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	18:41:51
Logística	CTM	OUTROS -	Bruna G	Maria C	GEOD	0:09:58

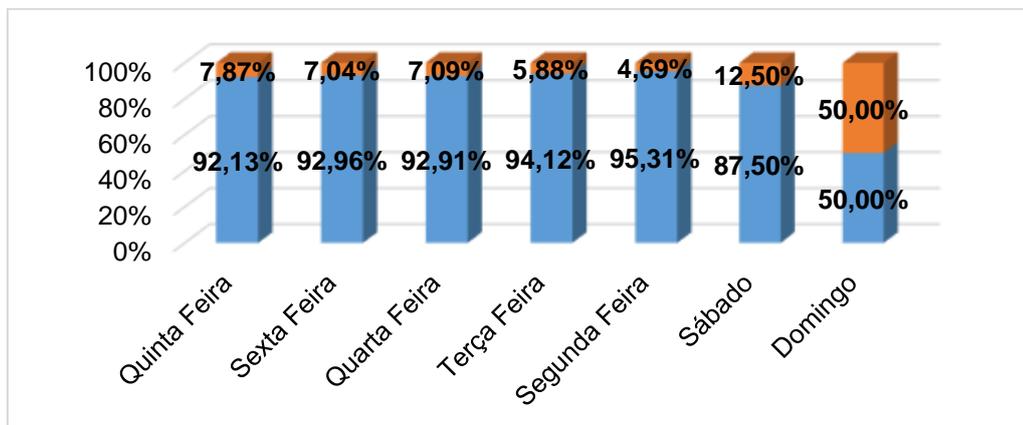
Fonte: Autor (2024)

A conclusão da análise das colunas de interesse assim como a percepção de grande variação do tempo de atendimento para situações semelhantes possibilitou o início da etapa engenharia de atributos.

#### 5.1.4 Engenharia de atributos

Com o intuito de verificar se há a violação de SLA em quantidade que chame a atenção em algum dia específico da semana foi criada uma coluna “Dia\_da\_semana”. Por meio do mesmo raciocínio utilizado para verificar o percentual de solicitações vencidas das colunas de interesse, não foi possível identificar dia para ser destacado salvo os dias de final de semana, conforme abaixo representado pelo gráfico 13.

Gráfico 13 - Coluna dias da semana



Fonte: Autor (2024)

Concluídas as etapas de análise exploratória de dados e engenharia de atributos, foi iniciada a próxima fase que consiste na preparação dos dados para que sejam processados pelos algoritmos de ML.

### 5.1.5 Pré-processamento

Concluído o comparativo percentual das colunas selecionadas entre solicitações atendidas e não atendidas dentro do prazo de atendimento e verificação se houve algo relevante em algum dia específico da semana foi criada a variável “solicitacoes\_resolvidas\_categorias” com as colunas de interesse para o objetivo da ferramenta. Sendo elas “Categoria”, “Subcategoria”, “Criado por”, “Tecnico”, “Local”, “Status do vencimento”, “Status da solicitação”, “Status da Primeira Resposta Atrasada”, “Dia\_da\_semana”.

Considerando que muitos algoritmos de ML exigem entradas numéricas e no estudo até então há somente variáveis categóricas, foi necessário realizar a codificação dessas variáveis categóricas em números por meio da técnica *Label Encoder* no Python. Para tanto foi importada a biblioteca *Sklearn.preprocessing* que possibilitou a transformação das variáveis da coluna Categoria em números. Temos como exemplo a seguir a Figura 27.

Figura 27 - Exemplo de uso do *Label encoder*



Fonte: Autor (2024)

A mesma técnica foi utilizada para as demais colunas de interesse e como resultado da etapa de pré-processamento temos o *dataset* exemplificado pela

Figura 28 conforme abaixo.

Figura 28 - Dataset para pré-processamento

```
In [181]: solicitacoes_resolvidas_categorias
```

```
Out[181]:
```

	Categoria	Subcategoria	Criado_por	Tecnico	Local	Status_do_vencimento	Status_da_Primeira_Resposta_Atrasada	Dia_da_semana
1	2	2	0	10	1	0	0	5
8	2	2	6	7	0	1	1	5
9	2	2	6	7	0	1	1	5
10	2	2	6	7	0	1	1	6
11	2	2	6	7	0	1	1	4
...	...	...	...	...	...	...	...	...
1906	0	5	12	10	13	0	0	6
1907	0	11	49	10	13	0	0	5
1908	0	5	49	10	13	0	0	4
1912	0	23	21	10	14	0	0	6
1913	0	25	6	15	14	0	0	5

1654 rows x 8 columns

Fonte: Autor (2024)

Com o *dataset* parametrizado por meio da técnica *Label Encoder* a função `train_test_split` da biblioteca `sklearn.model_selection` foi importada com o intuito de viabilizar a divisão dos dados em conjuntos de dados e teste.

Foram selecionadas as colunas “Categoria”, “Subcategoria”, “Criado\_por”, “Tecnico”, Local, “Status\_do\_vencimento”, “Status\_da\_Primeira\_Resposta\_Atrasada”, “Dia\_da\_semana” e atribuídas a variável “`colunas_features`” e atribuídas a variável X conforme Figura 29 abaixo.

Figura 29 - Atribuição das *features* a variável X

```
# Features (atributos)
X = solicitacoes_resolvidas_categorias[colunas_features]
```

Fonte: Autor (2024)

A variável alvo Status do Vencimento foi atribuída a Y conforme Figura 30

Figura 30 - Variável alvo

```
# Variável-alvo
y = solicitacoes_resolvidas_categorias['Status_do_vencimento']
```

Fonte: Autor (2024)

Foi feita a divisão dos dados em treino e teste com 80% para treinamento e 20% dos dados para teste conforme Figura 31.

Figura 31 - Divisão treino / teste

```
# Divisão dos dados em conjuntos de treinamento e teste (80% treinamento, 20% teste)
X_treino, X_teste, y_treino, y_teste = train_test_split(X, y, test_size=0.2, random_state=42)
```

Fonte: Autor (2024)

Foi feita a exibição dos conjuntos resultantes conforme Figura 32

Figura 32 - Exibição dos conjuntos resultantes

```
# Exibir as formas dos conjuntos resultantes
print("Forma do conjunto de treinamento (X):", X_treino.shape)
print("Forma do conjunto de teste (X):", X_teste.shape)
print("Forma do conjunto de treinamento (y):", y_treino.shape)
print("Forma do conjunto de teste (y):", y_teste.shape)

Forma do conjunto de treinamento (X): (1331, 7)
Forma do conjunto de teste (X): (333, 7)
Forma do conjunto de treinamento (y): (1331,)
Forma do conjunto de teste (y): (333,)
```

Fonte: Autor (2024)

A conclusão da etapa de pré-processamento possibilitou o primeiro teste de treinamento que se deu por meio de um modelo de regressão logística multivariada no qual foram obtidos os resultados conforme Figura 33.

Figura 33 - Resultado regressão logística

```
from sklearn.metrics import classification_report
print(classification_report(y_teste, pred01))
```

	precision	recall	f1-score	support
0	0.96	1.00	0.98	314
1	0.83	0.26	0.40	19

Fonte: Autor (2024)

Com o resultado é possível observar uma precisão de 96% das ocorrências de fato pertencem a classe 0 (solicitações atendidas no prazo). No caso da classe 1 (solicitações não atendidas dentro do prazo) o percentual de acerto foi 83%.

Figura 34 - Resultado matriz de confusão

```
from sklearn.metrics import confusion_matrix
conf_mat = confusion_matrix(y_teste, pred01)
print(conf_mat)
```

```
[[313  1]
 [ 14  5]]
```

Fonte: Autor (2024)

Pela Figura 34, que representa o resultado da Matriz de confusão, é possível perceber que do total de 327 solicitações, verdadeiros positivos somados aos falsos negativos. Temos 313 solicitações que de fato foram atendidas dentro do prazo, representando 95,6% de acertos

para verdadeiros positivos. É fato a considerar o baixo número de solicitações não atendidas dentro do prazo na base.

Com a perspectiva de prever o tempo de atendimento em função das variáveis independentes “Categoria”, “Subcategoria”, “Criado\_por”, “Tecnico”, “Local”, “Status\_da\_Primeira\_Resposta\_Atrasada” foi escolhido o modelo de Regressão Linear com avaliação por de Mínimos Quadrados Ordinários (OLS, *Ordinary Least Squares*).

Não foi possível obter resultado superior a 0,035, que sugere uma variabilidade de apenas 3,5% da variável dependente conforme mostra a Figura 35.

Figura 35 - Modelo de regressão logística

```
print(modelo_v1.summary())
```

OLS Regression Results

```
=====
```

Dep. Variable:	Tempo decorrido em segundos	R-squared:	0.035
----------------	-----------------------------	------------	-------

Fonte: Autor (2024)

Houve a tentativa de melhorar o resultado inserindo e removendo algumas variáveis na lista de atributos, conforme Figura 367.

Figura 36 - Variação das variáveis da lista features

```
# Defina as colunas que serão usadas como features (atributos)
#colunas_features_sem_criado_por = ['Categoria', 'Subcategoria', 'Tecnico', 'Local', 'Status_da_Primeira_Resposta_Atrasada']

# Defina as colunas que serão usadas como features (atributos)
#colunas_features_sem_resposta_atrasada = ['Categoria', 'Subcategoria', 'Criado_por', 'Tecnico', 'Local']

#colunas_features_sem_Tecnico = ['Categoria', 'Subcategoria', 'Criado_por', 'Local']
```

Fonte: Autor (2024)

Na aplicação dos algoritmos de ML, ainda que os resultados iniciais obtidos por uma análise de regressão multivariada logística a princípio apresentaram taxa de acerto de 95% para chamados atendidos dentro do prazo e 83% para chamados não atendidos dentro do prazo, a viabilidade de realização de predição de tempo com os dados existentes na organização não foi identificada.

Diversas outras configurações foram testadas com a mais diversa combinação em variáveis independentes considerando:

- Transformação de valores de variáveis independentes categóricas por valores em sequência de identificação;

- Transformação de valores de variáveis independentes categóricas por valores em ordem de frequência;
- Transformação de valores de variáveis independentes categóricas por valores em ordem de média de tempo;
- Transformação de valores de variáveis independentes categóricas por valores das médias de tempo de cada categoria;
- Transformação direta de variáveis independentes categóricas para zeros e uns.

Mesmo correndo risco de *overfitting* em algumas dessas situações não foram alcançados resultados superiores a 20% nas análises de regressão por R-quadrado.

Cortes na base de dados para tentar predições ao menos com atendimentos de menos de 50, 100 e 150 horas diminuindo o desvio padrão foram também alvo das análises e outros algoritmos como KNN e Naive Bayes foram testados sem sucesso.

Sendo assim, não foi possível realizar a predição de tempo a partir das variáveis independentes existentes na base de dados existente. Considera-se a possibilidade de análises com bases de dados maiores e ainda utilização de outras técnicas, o que não foi possível em tempo hábil para a finalização da entrega prevista para este trabalho, mas que podem ser consideradas como estudos futuros.

A reformulação da coleta de dados com a definição de outras variáveis independentes também pode ser sugerida para a maior assertividade na predição de tempo dos atendimentos.

#### 5.1.6 Avaliação do processo por profissionais especialistas

Após toda a realização das análises, o PRODIMAH e resultados obtidos foram apresentados a profissionais especialistas na área. Estes foram questionados sobre a factibilidade, utilidade e usabilidade. Os profissionais têm larga experiência na área.

As principais considerações foram quando questionados sobre:

a) Factibilidade:

O que achou da proposta do processo quanto a sua factibilidade (o processo proposto faz sentido para conhecer melhor o atendimento no *helpdesk* e eventuais melhorias no acordo a nível de serviços com o cliente - SLA)?

R1 – “Sim, é factual, pois analisa a base de dados e checa o desempenho passado, antes de se propor uma ou novas soluções para o SLA.”

R2 – “Sim faz sentido. A análise de dados a partir da base dá a possibilidade de entender tendências do passado e mudar para aprimorar.”

R3 – “A proposta apresenta-se plenamente factível para o conhecimento e melhor atendimento do *helpdesk*. Isto é possível pela disponibilidade dos dados de atendimentos que documentam os registros operacionais destes. E, considerando os possíveis resultados obtidos pela aplicação de algoritmos de ML para a definição do ‘tempo’, seria de grande justificativa para um estudo de revisão de processos e procedimentos existentes para a entrega e suporte dos nossos serviços, ‘identificando’ áreas cujos processos poderiam ser otimizados ou aprimorados para melhorar a eficiência e qualidade dos serviços, conforme citado no trabalho. Quanto a factibilidade de aplicação de algoritmos de ML para a melhor previsão do SLA projetado, a existência de dados operacionais necessários (categóricos, como descreve o autor), tais como ‘Departamento’, ‘Categoria’, ‘Subcategoria’, ‘Criado por’, ‘tecnico’, ‘Local’, ‘status da solicitação’, ‘Status do vencimento’, ‘Status da primeira resposta atrasada’, ‘Identificação da solicitação’, ‘Criado por’ e ‘Status do vencimento’ tornam este objetivo realizável, caso seja aplicado”

b) Utilidade:

O que achou da proposta quanto a utilidade (a proposta foi útil para vocês conhecerem melhor o atendimento no *helpdesk* e eventuais melhorias no acordo a nível de serviços com seus clientes - SLA)?

R1 – “É útil. Faz conhecer os desvios existentes e conectar com as dores que escutamos.”

R2 – “Útil pois o conhecimento dos atendimentos e tendências dá a possibilidade de verificar o que está errado e aplicar melhorias.”

R3 – “A proposta apresentou-se de extrema utilidade. Notamos que foi possível uma identificação do que denominamos de ‘chamadas fora do contexto de um possível SLA’, ou seja, chamadas de problemas que não são de âmbito do nosso contrato, como por exemplo, as identificadas como ‘Infraestrutura’, que correspondem a problemas técnicos do ambiente do cliente e que influenciam no desempenho, ou mesmo operação do nosso produto/sistema (resultado citado pelo autor “INFRA-ESTUTURA-NOVO com 45,16% das solicitações não atendidas dentro do prazo, ou Evsys (produto de outro fornecedor, contratado pelo cliente, e

que faz conexão interdependente com nosso produto). Outro exemplo quando identificado pelo nome do produto ‘CTMS’: pode indicar desconhecimento do que está acontecendo pelo entrante da chamada. Sendo uma descrição genérica, pode corresponder, além de uma falha que não se consegue identificar, também um desconhecimento do produto, tanto por parte do receptor do atendimento (problema de treinamento da nossa empresa), ou mesmo do emissor da chamada (problema de treinamento do cliente usuário de nosso produto). Outro fator que indica a usabilidade seria a possível análise preditiva do tempo de demora em não resolução, para uma melhor definição do SLA estipulado em nossos contratos. Como não utilizamos nenhuma análise neste sentido, o proposto de aplicação dos algoritmos de ML, obtidos por uma análise de regressão multivariada logística em princípio apresentaram, no nosso entendimento, uma interessante “taxa de acerto de 95% para chamados atendidos dentro do prazo e 83% para chamados não atendidos dentro do prazo” (conforme conclui o estudo).

c) Usabilidade:

O que achou da usabilidade proposta do processo quanto a sua usabilidade (o processo proposto é fácil de usar nas organizações para conhecer melhor o atendimento no helpdesk e identificar eventuais melhorias no acordo a nível de serviços com o cliente - SLA)?

R1 – “Sim, a divulgação e geração da expectativa dos prazos dos serviços disponíveis é essencial.”

R2 – “Sim, há a necessidade de disponibilidade de ambas as partes para alinhar as expectativas.”

R3 – “Quanto a sua usabilidade, no sentido de utilizar em outras organizações, consideramos sim de grande valor, porém ressaltando que este processo pode encontrar dificuldades no sentido de definição ou existência de variáveis ditas e citadas ‘categóricas’ neste estudo, pois elas podem variar de empresa ou da característica do produto, ou mesmo de um serviço que depende de um bem portador (produto) para ser executado. E lembrando que para melhor aprimoramento as empresas que se submeteriam a esta abordagem teriam de estar dispostas a mudança de processos e rotinas para melhor implementação do modelo. Ressaltamos isto porque muitas empresas utilizam sistema de *helpdesk* padronizados em produtos/pacotes oferecidos por fornecedores especializados, e que acabam exigindo uma padronização de processos, muitas vezes não condizente com a necessidade particular de atendimento de um cliente e seus necessários registros.”

### 5.1.7 *Explicitação das aprendizagens*

Temos um cenário empresarial contemporâneo que é marcado por uma rápida evolução, no qual empresas necessitam explorar novas tecnologias digitais, o que impacta direta e fortemente suas operações e conceitos de gestão. Essa rápida evolução da TD destaca a importância crescente dos dados. Nesse contexto, o GSTI surge como uma abordagem consistente para gerenciar a função de TI com a implementação de sistemas de *helpdesk*. Neste contexto destaca-se a importância, por exemplo, do atendimento dessas solicitações dentro do prazo, assim como a classificação precisa de tickets, o que viabiliza a reutilização eficiente de soluções, resultando em uma padronização e consequente redução no tempo de resolução e o que ocasiona o aumento da satisfação do cliente.

Pode-se verificar, em especial pela revisão bibliográfica, que estudos para predição do tempo de atendimento de chamados apresenta uma abordagem inovadora e promissora para aprimorar a eficiência dos serviços de *helpdesk* que inclusive possibilita a análise do sentimento dos usuários em relação ao serviço prestado.

Considerando a IA como resultado da TD, é notório que a sua evolução transcende da academia para o mercado comercial, sendo um recurso fundamental para a evolução tecnológica. O uso conjunto de Big Data, Ciência de dados, IA e ML emerge como recurso-chave com o destaque para IA que é utilizada nas mais variadas formas de negócio, desempenhando um papel crucial na resolução de diversos problemas de negócios.

Com a proposta para resolução do problema foi destacada a importância de compreender o contexto o qual o *helpdesk* está inserido, e enfatizada a necessidade de alinhamento com os objetivos estratégicos, estrutura organizacional e cultura da empresa. A reunião para obtenção de dados sobre atendimentos e análises inicial e exploratória dos mesmos foram ressaltadas como etapas cruciais para entender a situação atual e identificar tendências. A revisão de SLA e métricas de desempenho, juntamente com a otimização dos processos existentes, foram vistas como medidas essenciais para garantir a eficiência e qualidade dos serviços do *helpdesk*. Também foi destacada a importância de estabelecer canais de comunicação eficazes e definir responsabilidades claras para promover uma colaboração bem-sucedida entre a empresa e o provedor de serviços de TI, o que proporcionará um relacionamento de confiança ao longo do tempo.

No desenvolvimento do processo na macro etapa técnica não foi possível identificar dificuldade e esforço significativo na etapa de limpeza de dados conforme verificado na literatura o que foi justificado pelo *dataset* ser composto por uma única tabela. Recursos de processamento computacional mesclados com ferramentas estatísticas possibilitaram e viabilizaram identificar características que denotaram ser fundamentais e que demonstram o percentual no atendimento dos chamados dentro ou fora do prazo para cada uma das colunas selecionadas.

Durante a análise exploratória foi verificado que a categoria CTM absorveu grande parte do volume de dados em relação as demais categorias.

Vale ressaltar algumas observações na análise da distribuição do volume de chamados para a coluna Departamento onde percebe-se o grande volume de caso nos departamentos Logística, TI e Infraestrutura. Ainda considerando a coluna Departamento, foi observado a duplicidade de dois departamentos, sendo eles, “Administração” e “Administracion”. Da mesma forma foram encontrados os departamentos TI, Informática e Tecnologia da Informação. Por fim, também vale notar o departamento não atribuído com o montante de 134 casos.

No que tange à coluna Subcategoria as observações a serem considerada foram para o montante de 324 casos para “Outros” ocupando a segunda posição em quantidade absoluta dos casos assim como diversas subcategorias com o montante de casos inferior a 10.

Partindo para a análise da coluna “Criado por”, também foi encontrada significativa discrepância entre os valores permitindo notar a diferença de 245 solicitações criadas pelo analista mais representativo e apenas uma solicitação para o analista menos representativo em termos de volume.

Característica semelhante foi encontrada para a coluna Técnico onde pode ser percebido um montante de 411 solicitações resolvidas pelo analista com maior volume e apenas uma solicitação para os três analistas com menor volume de atendimento.

Na verificação da quantidade de casos por departamento foi identificado um desvio no processo de categorização dos departamentos com duas categorias para o departamento Administração o que sugere melhor definição do processo de gestão da base de dados.

A descoberta de grande variação no tempo de atendimento durante a análise exploratória mesmo ao considerar os mesmos departamentos, categorias, subcategorias, analistas responsáveis pela criação da solicitação, técnicos responsáveis pelo atendimento e locais, foi

constatada. Essa constatação demonstra a complexidade e influência de diversos fatores no processo de atendimento, destacando a importância de uma análise mais aprofundada para compreender as causas dessa variabilidade. O que foi aprendido na análise exploratória dos dados revela a necessidade de implementar medidas para reduzir essa variação e otimizar o tempo de atendimento, visando melhorar a eficiência e a qualidade dos serviços prestados pelo *helpdesk* o que possibilitará o uso de recursos avançados de inteligência artificial no futuro. Foi verificado na literatura que os algoritmos de ML analisam os dados em busca de padrões e, para o presente estudo, foi de suma importância a identificação da grande variabilidade do tempo de atendimento mesmo que para situações semelhantes.

A engenharia de atributos ofereceu uma abordagem que possibilitou criar uma coluna para representar os dias da semana permitindo verificar que não houve algum desvio para um dia específico salvo os finais de semana.

Na fase de pré-processamento a técnica *Label encoder* por meio da biblioteca `sklearn.preprocessing` possibilitou a transformação das variáveis categóricas em números o que viabilizou o processamento de algoritmos de ML.



## CONCLUSÃO

Ainda que tenha sido destacada a importância dos investimentos significativos em TI incluindo programas de transformação digital, com o intuito de aprimorar a qualidade e a entrega de serviços de *helpdesk*, foi evidenciado que o rápido desenvolvimento tecnológico desafia as organizações a adotarem estratégias mais criteriosas na escolha e implementação de tecnologias, garantindo alinhamento efetivo com os objetivos estratégicos. A compreensão profunda e a aplicação ponderada das transformações tecnológicas são cruciais para o sucesso empresarial em um ambiente cada vez mais digitalizado. Neste contexto este trabalho identificou a necessidade de adotar um processo que permita diagnosticar e aprimorar o serviço de *helpdesk*, imerso em um ambiente de transformações aceleradas. Para isso, a partir de uma revisão bibliográfica construiu-se a ferramenta de apoio PRODIMAH, Processo de Diagnóstico e Melhoria de Atendimento em *Helpdesk*. O processo proposto foi construído seguindo os passos do *DSR*. Nesse contexto foi proposto, desenvolvido, aplicado e avaliado.

O processo construído foi aplicado para a avaliação em uma empresa de capital 100% nacional, especializada em TI e Integradora de Soluções cujo principal foco é o fornecimento de soluções integradas em serviços de TI, softwares, aplicativos e sistemas de gestão empresarial que emprega as mais conceituadas plataformas globais para os setores de Agronegócio, Manufatura, Automotivo, Logística, Serviços e Varejo.

Após uma detalhada investigação sobre a base de dados disponibilizada pela empresa, usando estatística descritiva prevista no processo, apresentou-se um relatório da análise e diagnóstico dos serviços de *helpdesk* prestados com algumas sugestões de melhorias principalmente relacionadas a melhor coleta de dados dos serviços de *helpdesk*.

Na aplicação do processo para a sua avaliação, ainda que os resultados iniciais obtidos por uma análise de regressão multivariada logística a princípio apresentaram taxa de acerto de 95% para chamados atendidos dentro do prazo e 83% para chamados não atendidos dentro do prazo, a viabilidade de realização de predição de tempo com os dados existentes na organização não foi identificada. Após diversas tentativas usando as mais variadas técnicas tanto de acomodação da base de dados quanto de algoritmos de ML não se obteve a possibilidade de ajustes do SLA a partir das predições de tempo.

Mesmo assim, a apresentação do diagnóstico que faz parte dos resultados para a organização serviu para aprimorar, a partir do autoconhecimento, a prestação de serviços.

O diagnóstico da situação atual, viabilizado pelo PRODIMAH, proposta de solução apresentada, considerando tanto os níveis de serviço quanto os recursos utilizados, permitirá a definição do SLA com um detalhamento abrangente sobre os objetivos e as datas de revisão dos serviços. Isso possibilitará a adaptação dos escopos de atendimento, proporcionando uma visibilidade clara dos níveis de serviço e do volume da demanda a ser atendida. Portanto, é evidente que o uso do processo para diagnóstico e melhoria de atendimento em *helpdesk* possibilitará o melhor planejamento do SLA, com a identificação precisa dos objetivos e requisitos, juntamente com a designação clara dos responsáveis pelos serviços prestados, sendo este essencial para possibilitar o uso de recursos de inteligência artificial.

Por fim o PRODIMAH e os resultados obtidos com sua aplicação foram apresentados a especialistas que deram sua opinião sobre o processo proposto quanto a sua factibilidade, utilidade e usabilidade.

Com relação à factibilidade foi dito que o processo é factível e se destacou a importância da análise da base de dados para identificar tendências, checar desempenho passado e propor novas soluções para o SLA.

Com relação à utilidade se destacou o conhecer os desvios existentes e conectar com as dores apontadas e verificar o que está errado e aplicar melhorias.

Com relação à usabilidade se destacou a importância da divulgação e geração da expectativa dos prazos dos serviços disponíveis e se alertou para a necessidade da disponibilidade de ambas as partes do atendimento alinhar as expectativas.

Este estudo focou na análise da base de dados de uma empresa específica. Futuras pesquisas podem explorar diferentes bases de dados para validar a generalização dos resultados. Além disso, a base disponibilizada foi utilizada com ferramentas como Python e Excel, sugerindo a possibilidade de investigar outras ferramentas de análise de dados. Novos estudos também podem considerar a inclusão de uma variedade maior de categorias de interesse na base de dados, permitindo a transformação para tipos de dados compatíveis com uma gama mais ampla de algoritmos de ML. Este estudo não aplicou técnicas de normalização e padronização, indicando uma área promissora para pesquisas futuras que poderiam avaliar o impacto dessas técnicas na performance dos modelos. Na regressão logística, a coluna “precision” foi utilizada como parâmetro de viabilidade, mas estudos futuros poderiam experimentar outros parâmetros de avaliação. Similarmente, na regressão linear, o parâmetro de viabilidade foi o R-Squared, o que abre a possibilidade de novos estudos que considerem diferentes métricas de desempenho.

## REFERÊNCIAS

- AHMED, S. *et al.* **An Empirical Analysis of State-of-Art Classification Models in an IT Incident Severity Prediction Framework.** Applied Sciences, v. 13, n. 6, p. 3843, 2023.
- AKKIRAJU, R. *et al.* **Characterizing machine learning processes: A maturity framework.** Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v. 12168 LNCS, p. 17–31, 2020.
- AL-ASHMOERY, Y. *et al.* **Impact of IT Service Management and ITIL Framework on the Businesses.** International Conference of Modern Trends in ICT Industry: Towards the Excellence in the ICT Industries, MTICTI 2021, n. January 2022, 2021.
- ALEXAN, A. *et al.* **Analysis of activity detection data pre-processing.** SIITME 2019 - 2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging, Proceedings, n. October, p. 282–286, 2019.
- AL-HAWARI, F.; BARHAM, H. **A machine learning based help desk system for IT service management.** Journal of King Saud University - Computer and Information Sciences, v. 33, n. 6, p. 702–718, 2021.
- ALZUBAIDI, A. *et al.* **Blockchain-Based SLA Management in the Context of IoT.** IT Professional, v. 21, n. 4, p. 33–40, 2019.
- ANDREWS, A. A.; BEAVER, P.; LUCENTE, J. **Towards better help desk planning: Predicting incidents and required effort.** Journal of Systems and Software, v. 117, p. 426–449, 2016.
- BARCZEWSKI, A.; BEZERIANOS, A.; BOUKHELIFA, N. **How domain experts structure their exploratory data analysis: Towards a machine-learned storyline.** Conference on Human Factors in Computing Systems - Proceedings, p. 1–8, 2020.
- BEKAR, E. T.; NYQVIST, P.; SKOOGH, A. **An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study.** Advances in Mechanical Engineering, v. 12, n. 5, p. 1–14, 2020.
- BELLO-ORGAZ, G.; JUNG, J. J.; CAMACHO, D. **Social big data: Recent achievements and new challenges.** Information Fusion, v. 28, p. 45–59, 2016.
- BERRAL, J. L. *et al.* **Towards energy-aware scheduling in data centers using machine learning.** Proceedings of the e-Energy 2010 - 1st Int'l Conf. on Energy-Efficient Computing and Networking, v. 2, p. 215–224, 2010.
- BEZERRA, A. *et al.* **Extracting value from industrial alarms and events: A data-driven approach based on exploratory data analysis.** Sensors (Switzerland), v. 19, n. 12, 2019.
- BHATTACHARYA, P. **Artificial Intelligence in the Boardroom: Enabling “Machines” to “Learn” to Make Strategic Business Decisions.** ITT 2018 - Information Technology Trends: Emerging Technologies for Artificial Intelligence, p. 170–174, 2019.
- BILALLI, B. *et al.* **PRESISTANT: Learning based assistant for data pre-processing.** Data and Knowledge Engineering, v. 123, n. July, p. 101727, 2019.
- BOIRE, R. **Artificial intelligence(AI), automation, and its impact on data science.** Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017, v. 2018- Janua, p. 3571–3574, 2017.

CAVLAK, N.; COP, R. **The role of big data in digital marketing**. *Advanced Digital Marketing Strategies in a Data-Driven Era*, v. 1, n. 1, p. 16–33, 2021.

CHEN, X. *et al.* **Neural feature search: A neural architecture for automated feature engineering**. *Proceedings - IEEE International Conference on Data Mining, ICDM*, v. 2019-Novem, n. Icdm, p. 71–80, 2019.

CYBULSKI, J. L.; SCHEEPERS, R. **Data science in organizations: Conceptualizing its breakthroughs and blind spots**. *Journal of Information Technology*, v. 36, n. 2, p. 154–175, 2021.

DAELE, S. VAN; JANSSENSWILLEN, G. **Identifying the Steps in an Exploratory Data Analysis: A Process-Oriented Approach**. [s.l.] Springer Nature Switzerland, 2023. v. 468 LNBIP

DAVIS, J. J.; FOO, E. **Automated feature engineering for HTTP tunnel detection**. *Computers and Security*, v. 59, p. 166–185, 2016.

DEMCHENKO, Y. *et al.* **EDISON data science framework: A foundation for building data science profession for research and industry**. *Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom*, v. 0, n. February 2017, p. 620–626, 2016.

SANTOS, H. G. *et al.* **Machine learning for predictive analyses in health: An example of an application to predict death in the elderly in São Paulo, Brazil**. *Cadernos de Saude Publica*, v. 35, n. 7, 2019.

EICKEMEYER, S. C. *et al.* **Acting instead of reacting—ensuring employee retention during successful introduction of i4.0**. *Applied System Innovation*, v. 4, n. 4, p. 1–18, 2021.

ELREFAI, A. T.; ELGAZZAR, M. H.; KHODEIR, A. N. **Using Artificial Intelligence in Enhancing Banking Services**. *2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC 2021*, p. 980–986, 2021.

Forconi, M. G.; **Estratégias de melhorias do SLA ( Service Level Agreement) através da criação de um documento de oferta básica de uma empresa**. *Universiade Federal do Paraná*. Curitiba. 2016.

GANDOMI, A.; HAIDER, M. **Beyond the hype: Big data concepts, methods, and analytics**. *International Journal of Information Management*, v. 35, n. 2, p. 137–144, 2015.

GARVEY, C. M.; JONES, R. **Is There a Place for Theoretical Frameworks in Qualitative Research?** *International Journal of Qualitative Methods*, v. 20, p. 1–7, 2021.

GHOBAKHLOO, M.; IRANMANESH, M. **Digital transformation success under Industry 4.0: a strategic guideline for manufacturing SMEs**. *Journal of Manufacturing Technology Management*, v. 32, n. 8, p. 1533–1556, 2021.

GOMES, S. B.; FALBO, R. DE A.; MENEZES, C. S. DE. **Um Modelo para Acordo de Nível de Serviço em TI**. p. 191–205, 2021.

GREGOR, S.; HEVNER, A. R. **Positioning and presenting design science research for maximum impact**. *MIS Quarterly: Management Information Systems*, v. 37, n. 2, p. 337–355, 2013.

GROVER, V. *et al.* **Creating Strategic Business Value from Big Data Analytics: A Research Framework**. *Journal of Management Information Systems*, v. 35, n. 2, p. 388–423, 2018.

GUEDES, L. G. DE R. *et al.* **Service Level Agreements for Management Information Systems in Business and Governmental Organizations / Acordos de Nível de Serviço para Sistemas de Informação de Gestão em Organizações Comerciais e Governamentais**. *Brazilian Journal of Business*, v. 3, n. 3, p. 2489–2501, 2021.

- GUIMARÃES, L. M. S.; MEIRELES, M. R. G.; DE ALMEIDA, P. E. M. **Evaluation of the preprocessing and training stages in text classification algorithms in the context of information retrieval.** *Perspectivas em Ciencia da Informacao*, v. 24, n. 1, p. 169–190, 2019.
- GUNAY, M. *et al.* **Digital Data Forgetting: A Machine Learning Approach.** ISMSIT 2018 - 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings, p. 0–3, 2018.
- GUPTA, D.; RANI, R. **A study of big data evolution and research challenges.** *Journal of Information Science*, v. 45, n. 3, p. 322–340, 2019.
- HORTA, R. A. M.; ALVES, F. J. DOS S.; CARVALHO, F. A. A. DE. **Seleção de atributos na previsão de insolvência: aplicação e avaliação usando dados brasileiros recentes.** *RAM. Revista de Administração Mackenzie*, v. 15, n. 1, p. 125–151, 2014.
- ILYAS, I. F.; REKATSINAS, T. **Machine Learning and Data Cleaning: Which Serves the Other?** *Journal of Data and Information Quality*, v. 14, n. 3, 2022.
- JURJ, D. I. *et al.* **Analysis of Data Cleaning Techniques for Electrical Energy Consumption of a Public Building.** UPEC 2020 - 2020 55th International Universities Power Engineering Conference, Proceedings, 2020.
- KAYNAK, O. **The golden age of Artificial Intelligence.** *Discover Artificial Intelligence*, v. 1, n. 1, dez. 2021.
- KEWO, A.; MANEMBU, P.; NIELSEN, P. S. **Data pre-processing techniques in the regional emission's load profiles case.** 2019 6th International Conference on Control, Decision and Information Technologies, CoDIT 2019, p. 2016–2021, 2019.
- KHAN, M. A. U. D.; UDDIN, M. F.; GUPTA, N. **Seven V's of Big Data understanding Big Data to extract value.** Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education - "Engineering Education: Industry Involvement and Interdisciplinary Trends", ASEE Zone 1 2014, 2014.
- KHAN, S. U. *et al.* **Adaptive Runtime Monitoring of Service Level Agreement Violations in Cloud Computing.** *Computers, Materials and Continua*, v. 71, n. 2, p. 4199–4220, 2022.
- KOSTAKIS, P.; KARGAS, A. **Big-data management: A driver for digital transformation?** *Information (Switzerland)*, v. 12, n. 10, 2021.
- KRISHNAN, S. *et al.* **ActiveClean: An interactive data cleaning framework for modern machine learning.** Proceedings of the ACM SIGMOD International Conference on Management of Data, v. 26-June-20, p. 2117–2120, 2016.
- LANGER, T.; MEISEN, T. **Towards utilizing domain expertise for exploratory data analysis.** *ACM International Conference Proceeding Series*, 2019.
- LEE, C. Y.; CHIEN, C. F. **Pitfalls and protocols of data science in manufacturing practice.** *Journal of Intelligent Manufacturing*, v. 33, n. 5, p. 1189–1207, 2022.
- LEE, I. **Big data : Dimensions , evolution , impacts , and challenges.** *Business Horizons*, n. 2017, 2020.
- LENGAUER, T. **Statistical Data Analysis in the Era of Big Data.** n. 7, p. 831–841, 2020.
- LU, J.; CAIRNS, L.; SMITH, L. **Data science in the business environment: customer analytics case studies in SMEs.** *Journal of Modelling in Management*, v. 16, n. 2, p. 689–713, 2020.

- MACLEAN, D.; TITAH, R. **Implementation and impacts of IT Service Management in the IT function.** International Journal of Information Management, v. 70, p. 102628, 1 jun. 2023.
- MALIK, H.; CHAUDHARY, G.; SRIVASTAVA, S. **Digital transformation through advances in artificial intelligence and machine learning.** Journal of Intelligent and Fuzzy Systems, v. 42, n. 2, p. 615–622, 2022.
- MAO, H.; ZHANG, T.; TANG, Q. **Research framework for determining how artificial intelligence enables information technology service management for business model resilience.** Sustainability (Switzerland), v. 13, n. 20, 2021.
- MATT, C.; HESS, T.; BENLIAN, A. **Digital Transformation Strategies.** Business and Information Systems Engineering Gabler Verlag, 1 out. 2015.
- MENDONÇA, F. M.; DANTAS, M. A. R. **Covid-19: Where is the Digital Transformation, Big Data, Artificial Intelligence and Data Analytics?** Revista do Serviço Público, v. 71, p. 212–234, 2020.
- MIGUEL, P. A. C. **Estudo de caso na engenharia de produção: estruturação e recomendações para sua condução.** Production, v. 17, n. 1, p. 216–229, 2007.
- MIKLOSIK, A. *et al.* **Towards the Adoption of Machine Learning-Based Analytical Tools in Digital Marketing.** IEEE Access, v. 7, p. 85705–85718, 2019.
- MUGGE, P. *et al.* Patterns of Digitization: A Practical Guide to Digital Transformation. **Research Technology Management**, v. 63, n. 2, p. 27–35, 2020.
- MUSUMECI, F. *et al.* **A Tutorial on Machine Learning for Failure Management in Optical Networks.** Journal of Lightwave Technology, v. 37, n. 16, p. 4125–4139, 2019.
- OSTBLOM, J.; TIMBERS, T. **Opinionated Practices for Teaching Reproducibility: Motivation, Guided Instruction and Practice.** Journal of Statistics and Data Science Education, v. 30, n. 3, p. 241–250, 2022.
- PAHLEVAN SHARIF, S.; MURA, P.; WIJESINGHE, S. N. R. **Systematic Reviews in Asia: Introducing the “PRISMA” Protocol to Tourism and Hospitality Scholars.** Perspectives on Asian Tourism, v. Part F184, p. 13–33, 2019.
- Pereira, A.S.; Manosso, T.W.S.; Fossatti, E.C.; Berti, S.M.; **Regressão Linear Múltipla Como simplificar por meio do Excel e SPSS?.** Universidade de Passo Fundo- Faculdade de Ciências Econômicas, Administrativas e Contábeis Centro de Pesquisa e extensão. FEAC.2019
- PEOPLES, C. *et al.* **Analysis of a Personalized Provision of Service Level Agreement (SLA) Algorithm.** Electronics (Switzerland), v. 12, n. 5, p. 1–18, 2023.
- PREEZ, A. DU; OOSTHUIZEN, G. A. **Machine learning in cutting processes as enabler for smart sustainable manufacturing.** Procedia Manufacturing, v. 33, p. 810–817, 2019.
- PRIESTLEY, J.; MCGRATH, R. J. **The evolution of data science: A new mode of knowledge production.** International Journal of Knowledge Management, v. 15, n. 2, p. 97–109, 2019.
- REZIG, E. K. *et al.* **Data civilizer 2.0: A holistic framework for data preparation and analytics.** Proceedings of the VLDB Endowment, v. 12, n. 12, p. 1954–1957, 2018.
- RODRÍGUEZ GALLARDO, J. A.; LÓPEZ DE LA MADRID, M. C.; ESPINOZA DE LOS MONTEROS CÁRDENAS, A. **Study of the implementation of Help Desk software in an institution of higher education.** PAAKAT: Revista de Tecnología y Sociedad, v. 8, n. 14, p. 1–20, 2018.

ROSSI, E.; RUBATTINO, C.; VISCUSI, G. **Big data use and challenges: Insights from two internet-mediated surveys.** *Computers*, v. 8, n. 4, p. 1–15, 2019.

SALTZ, J.; SHAMSHURIM, I.; CONNORS, C. **Predicting Data Science Sociotechnical Execution Challenges by Categorizing Data Science Projects.** *Journal of the American Society for Information Science and Technology*, v. 64, n. July, p. 1852–1863, 2013.

SARWAR, M. I. *et al.* **Digital Transformation of Public Sector Governance With IT Service Management-A Pilot Study.** *IEEE Access*, v. 11, n. January, p. 6490–6512, 2023.

SELIM, A. **Systematic Review of Big Data, Digital Transformation Areas and Industry.** *Vision International Refereed Scientific Journal*, v. 6, n. 2, p. 27–41, 2021.

SHANMUGALINGAM, K. *et al.* **Corporate IT-Support Help-Desk Process Hybrid-Automation Solution with Machine Learning Approach.** 2019 *Digital Image Computing: Techniques and Applications, DICTA 2019*, n. MI, p. 1–7, 2019.

SILVA, R. B. **A Importância Do Sla No Gerenciamento Dos Serviços.** *X Congresso Nacional De Excelência Em Gestão*, p. 11, 2014.

SIRADJUDDIN, I. A.; SAKINAH, A.; SOPHAN, M. K. **Combination of feature engineering and feature learning approaches for classification on visual complexity images.** *International Journal of Innovative Computing, Information and Control*, v. 17, n. 3, p. 991–1005, 2021.

SONG, I. Y.; ZHU, Y. **Big data and data science: what should we teach?** *Expert Systems*, v. 33, n. 4, p. 364–373, 2016.

STEINBACHER, J. **A Hybrid Machine Learning Model for Efficient Classification of IT Support** *A Hybrid Machine Learning Model for Efficient Classification of.* n. January, 2023.

TITU, A. M.; STANCIU, A. **Acquiring business intelligence through data science: A practical approach.** *Proceedings of the 12th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2020*, n. MI, 2020.

UPADHYAY, K. *et al.* **Can't Understand SLAs? Use the Smart Contract.** *Proceedings - 2021 3rd IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2021*, p. 129–136, 2021.

VIEIRA, V. A. **As tipologias , variações e características da pesquisa de marketing.** *Revista da FAE*, v. 5, p. 61–70, 2002.

WANG, M.; DING, Z.; PAN, M. **LbR: A New Regression Architecture for Automated Feature Engineering.** *IEEE International Conference on Data Mining Workshops, ICDMW*, v. 2020- Novem, p. 432–439, 2020.

ZENG, X. *et al.* **Detection of SLA Violation for Big Data Analytics Applications in Cloud.** *IEEE Transactions on Computers*, v. 70, n. 5, p. 746–758, 2021.

ZUEV, D.; KALISTRATOV, A.; ZUEV, A. **Machine Learning in IT Service Management.** *Procedia Computer Science*, v. 145, p. 675–679, 2018.

## APÊNDICE

### CÓDIGO USADO PARA REALIZAÇÃO DAS ANÁLISES E ML

```

# Versão da Linguagem Python
from platform import python_version
print('Versão da Linguagem Python Usada Neste Jupyter Notebook:',
python_version())

# Imports
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")

# Versões dos pacotes usados neste jupyter notebook
%reload_ext watermark
%watermark -a "William Emidio de Noronha" --iversions

df = pd.read_csv("1683040374933.csv", header=5, delimiter=',', usecols=[
    'Identificação da solicitação'
    , 'Departamento'
    , 'Categoria'
    , 'Subcategoria'
    , 'Assunto'
    , 'Criado por'
    , 'Técnico'
    , 'Descrição'
    , 'Local'
    , 'Status da solicitação'
    , 'Status pendente'
    , 'Hora de criação'
    , 'Data de resposta'
    , 'Hora do vencimento'
    , 'Hora de conclusão'
    , 'Tempo decorrido'
    , 'Status do vencimento'
    , 'Resposta Devida às'
    , 'Status da Primeira Resposta Atrasada'
    , 'Horário da última atualização'
    , 'Hora da solução'
] )

df

df.size

```

```
df.shape

df.describe()

df.dtypes.value_counts()

df.info()

type(df)

# Limpeza dos dados

df.isnull().sum()

df.dropna().head()

df['Hora de criação'] = pd.to_datetime(df['Hora de criação'],
format='%d/%m/%Y %I:%M %p', errors='coerce')

df.info()

# Adicionar uma nova coluna para o dia da semana
df['Dia_da_semana'] = df['Hora de criação'].dt.day_name()

#df.drop('Dia da semana', axis=1, inplace=True)

df

df.dropna()

df.info()

df.duplicated().sum()

# Identificar linhas duplicadas
linhas_duplicadas = df[df.duplicated()]

# Mostrar as linhas duplicadas
print("Linhas Duplicadas:")
print(linhas_duplicadas)

# Remover linhas duplicadas do DataFrame original
df_sem_duplicatas = df.drop_duplicates()

# Mostrar informações sobre as duplicatas removidas
print("Número de linhas duplicadas removidas:", df.shape[0] -
df_sem_duplicatas.shape[0])
```

```

df_sem_duplicatas

# Remover a primeira linha do DataFrame
df = df.drop(df.index[0])

# Resetar o índice se necessário
df = df.reset_index(drop=True)

df_sem_duplicatas['Departamento'].value_counts()

import matplotlib.pyplot as plt

# Contagem de valores na coluna 'Departamento'
contagem_departamentos = df_sem_duplicatas['Departamento'].value_counts()

# Criar um gráfico de barras
ax = contagem_departamentos.plot(kind='bar')

# Adicionar rótulos e título
plt.xlabel('Departamento')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Departamento')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_departamentos):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

df_sem_duplicatas['Categoria'].value_counts()

# Contagem de valores na coluna 'Departamento'
contagem_departamentos = df_sem_duplicatas['Categoria'].value_counts()

# Criar um gráfico de barras
ax = contagem_departamentos.plot(kind='bar')

# Adicionar rótulos e título
plt.xlabel('Categoria')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Categoria')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_departamentos):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

df_sem_duplicatas['Subcategoria'].value_counts()

```

```

# Contagem de valores na coluna 'Subcategoria'
contagem_subcategorias = df_sem_duplicatas['Subcategoria'].value_counts()

# Criar um gráfico de barras com largura ajustada
ax = contagem_subcategorias.plot(kind='bar', width=0.5, figsize=(12, 8))
# Ajuste a largura conforme necessário

# Adicionar rótulos e título
plt.xlabel('Subcategoria')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Subcategoria')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_subcategorias):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

df_sem_duplicatas['Criado por'].value_counts()

# Contagem de valores na coluna 'Criado por'
contagem_subcategorias = df_sem_duplicatas['Criado por'].value_counts()

# Criar um gráfico de barras com largura ajustada
ax = contagem_subcategorias.plot(kind='bar', width=0.5, figsize=(15, 10))
# Ajuste a largura conforme necessário

# Adicionar rótulos e título
plt.xlabel('Criado por')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Criado por')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_subcategorias):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

df_sem_duplicatas['Técnico'].value_counts()

# Contagem de valores na coluna 'Técnico'
contagem_subcategorias = df_sem_duplicatas['Técnico'].value_counts()

# Criar um gráfico de barras com largura ajustada
ax = contagem_subcategorias.plot(kind='bar', width=0.5, figsize=(15, 10))
# Ajuste a largura conforme necessário

# Adicionar rótulos e título
plt.xlabel('Técnico')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Técnico')

```

```

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_subcategorias):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

df_sem_duplicatas['Local'].value_counts()

# Contagem de valores na coluna 'Local'
contagem_subcategorias = df_sem_duplicatas['Local'].value_counts()

# Criar um gráfico de barras com largura ajustada
ax = contagem_subcategorias.plot(kind='bar', width=0.5, figsize=(15, 10))
# Ajuste a largura conforme necessário

# Adicionar rótulos e título
plt.xlabel('Local')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Local')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_subcategorias):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

df_sem_duplicatas['Status do vencimento'].value_counts()

# Contagem de valores na coluna 'Status do vencimento'
contagem_subcategorias = df_sem_duplicatas['Status do
vencimento'].value_counts()

# Criar um gráfico de barras com largura ajustada
ax = contagem_subcategorias.plot(kind='bar', width=0.5, figsize=(15, 10))
# Ajuste a largura conforme necessário

# Adicionar rótulos e título
plt.xlabel('Status do vencimento')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Status do vencimento')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_subcategorias):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

df_sem_duplicatas['Status da Primeira Resposta Atrasada'].value_counts()

# Contagem de valores na coluna 'Status da Primeira Resposta Atrasada'

```

```

contagem_subcategorias = df_sem_duplicatas['Status da Primeira Resposta
Atrasada'].value_counts()

# Criar um gráfico de barras com largura ajustada
ax = contagem_subcategorias.plot(kind='bar', width=0.5, figsize=(10,7))
# Ajuste a largura conforme necessário

# Adicionar rótulos e título
plt.xlabel('Status da Primeira Resposta Atrasada')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Status da Primeira Resposta Atrasada')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_subcategorias):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

# Contagem de valores na coluna 'Status da Primeira Resposta Atrasada'
contagem_subcategorias = df_sem_duplicatas['Status do
vencimento'].value_counts()

# Criar um gráfico de barras com largura ajustada
ax = contagem_subcategorias.plot(kind='bar', width=0.5, figsize=(10,7))
# Ajuste a largura conforme necessário

# Adicionar rótulos e título
plt.xlabel('Status do vencimento')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Status do vencimento')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_subcategorias):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

# Mostrar o gráfico
plt.show()

df_sem_duplicatas['Dia_da_semana'].value_counts()

# Contagem de valores na coluna 'Dia_da_semana'
contagem_departamentos = df_sem_duplicatas['Dia_da_semana'].value_counts()

# Criar um gráfico de barras
ax = contagem_departamentos.plot(kind='bar')

# Adicionar rótulos e título
plt.xlabel('Dia_da_semana')
plt.ylabel('Número de Solicitações')
plt.title('Distribuição por Dia da semana')

# Adicionar rótulos nas barras
for i, valor in enumerate(contagem_departamentos):
    ax.text(i, valor + 0.1, str(valor), ha='center', va='bottom')

```

```

# Mostrar o gráfico
plt.show()

df_sem_duplicatas.isna().sum()

df_sem_duplicatas['Categoria'].isnull().sum()

# Separando os casos resolvidos
solicitacoes_resolvidas_categorias =
df_sem_duplicatas[df_sem_duplicatas['Status da solicitação'] ==
'Resolvido'][['Categoria', 'Subcategoria', 'Criado por', 'Técnico',
'Local', 'Status do vencimento', 'Status da solicitação', 'Dia_da_semana']]

# Filtrar apenas as solicitações resolvidas
solicitacoes_resolvidas = df_sem_duplicatas[df_sem_duplicatas['Status da
solicitação'] == 'Resolvido']

# Selecionar colunas específicas
solicitacoes_resolvidas_categorias = solicitacoes_resolvidas[['Status do
vencimento', 'Status da solicitação']]

# Mostrar as solicitações resolvidas e as colunas selecionadas
print(solicitacoes_resolvidas_categorias)

#Confirmando que temos só os resolvidos

solicitacoes_resolvidas["Status da solicitação"].describe()

solicitacoes_resolvidas_categorias =
df_sem_duplicatas[df_sem_duplicatas['Status da solicitação'] ==
'Resolvido'][['Categoria', 'Subcategoria', 'Criado por', 'Técnico',
'Local', 'Status do vencimento', 'Status da solicitação', 'Dia_da_semana']]

solicitacoes_resolvidas_categorias

solicitacoes_resolvidas_categorias.describe()

df_sem_duplicatas.shape

df_sem_duplicatas.isna().sum()

# Lista de colunas categóricas
cats = ['Departamento', 'Categoria', 'Subcategoria', 'Criado por',
'Técnico', 'Local', 'Status da Primeira Resposta Atrasada',
'Dia_da_semana']

cats

```

```

chart = sns.countplot(data=df_sem_duplicatas, x='Status do vencimento',
hue='Status do vencimento', palette=["red", "green"])

# Calcule e adicione os valores totais nas colunas
total_counts = df['Status do vencimento'].value_counts()
for i, p in enumerate(chart.patches):
    total = total_counts[i % len(total_counts)]
    chart.annotate(f'{total}', (p.get_x() + p.get_width() / 2.,
p.get_height()), ha='center', va='bottom', fontsize=12, color='black',
xytext=(0, 5), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Status do vencimento')
plt.ylabel('Contagem')
plt.title('Contagem de Status do Vencimento')

# Exiba o gráfico
plt.show()

plt.figure(figsize=(17,4))
# Count Plot da variável target
ax = sns.countplot(data=df_sem_duplicatas, x='Departamento', hue='Status do
vencimento', orient='v')

# Inclua rotação nos rótulos do eixo x (categorias)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90)

# Adicione rótulos aos topos das barras
for p in ax.patches:
    ax.annotate(f'{p.get_height()}', (p.get_x() + p.get_width() / 2.,
p.get_height()), ha='center', va='center', fontsize=12, color='black',
xytext=(0, 5), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Departamento')
plt.ylabel('Contagem')
plt.title('Contagem de Status do Vencimento por Departamento')

# Exiba o gráfico
plt.show()

plt.figure(figsize=(17,4))
# Count Plot da variável target
ax = sns.countplot(data=df_sem_duplicatas, x='Categoria', hue='Status do
vencimento', orient='v')

# Inclua rotação nos rótulos do eixo x (categorias)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90)

# Adicione rótulos aos topos das barras
for p in ax.patches:
    ax.annotate(f'{p.get_height()}', (p.get_x() + p.get_width() / 2.,
p.get_height()), ha='center', va='center', fontsize=12, color='black',
xytext=(0, 5), textcoords='offset points')

# Adicione rótulos aos eixos

```

```

plt.xlabel('Categoria')
plt.ylabel('Contagem')
plt.title('Contagem de Status do Vencimento por Categoria')

# Exiba o gráfico
plt.show()

plt.figure(figsize=(17, 8)) # Ajuste o tamanho conforme necessário

# Count Plot da variável target com barras na horizontal
ax = sns.countplot(data=df_sem_duplicatas, y='Subcategoria', hue='Status do
vencimento', orient='h')

# Adicione rótulos aos topos das barras
for p in ax.patches:
    ax.annotate(f'{p.get_width()}', (p.get_width(), p.get_y() +
p.get_height() / 2.), va='center', fontsize=12, color='black', xytext=(5,
0), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Contagem')
plt.ylabel('Subcategoria')
plt.title('Contagem de Status do Vencimento por Subcategoria')

# Exiba o gráfico
plt.show()

plt.figure(figsize=(17, 8)) # Ajuste o tamanho conforme necessário

# Count Plot da variável target com barras na horizontal
ax = sns.countplot(data=df_sem_duplicatas, y='Criado por', hue='Status do
vencimento', orient='h')

# Adicione rótulos aos topos das barras
for p in ax.patches:
    ax.annotate(f'{p.get_width()}', (p.get_width(), p.get_y() +
p.get_height() / 2.), va='center', fontsize=12, color='black', xytext=(5,
0), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Contagem')
plt.ylabel('Criado por')
plt.title('Contagem de Status do Vencimento por Criado por')

# Exiba o gráfico
plt.show()

plt.figure(figsize=(17, 8)) # Ajuste o tamanho conforme necessário

# Count Plot da variável target com barras na horizontal
ax = sns.countplot(data=df_sem_duplicatas, y='Técnico', hue='Status do
vencimento', orient='h')

# Adicione rótulos aos topos das barras
for p in ax.patches:

```

```

    ax.annotate(f'{p.get_width()}', (p.get_width(), p.get_y() +
p.get_height() / 2.), va='center', fontsize=12, color='black', xytext=(5,
0), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Contagem')
plt.ylabel('Técnico')
plt.title('Contagem de Status do Vencimento por Técnico')

# Exiba o gráfico
plt.show()

plt.figure(figsize=(17, 8)) # Ajuste o tamanho conforme necessário

# Count Plot da variável target com barras na horizontal
ax = sns.countplot(data=df_sem_duplicatas, y='Local', hue='Status do
vencimento', orient='h')

# Adicione rótulos aos topos das barras
for p in ax.patches:
    ax.annotate(f'{p.get_width()}', (p.get_width(), p.get_y() +
p.get_height() / 2.), va='center', fontsize=12, color='black', xytext=(5,
0), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Contagem')
plt.ylabel('Local')
plt.title('Contagem de Status do Vencimento por Local')

# Exiba o gráfico
plt.show()

# a análise abaixo é inconclusiva pois não é possível saber se o status
pendente é momentaneo

plt.figure(figsize=(17, 8)) # Ajuste o tamanho conforme necessário

# Count Plot da variável target com barras na horizontal
ax = sns.countplot(data=df_sem_duplicatas, y='Status pendente', hue='Status
do vencimento', orient='h')

# Adicione rótulos aos topos das barras
for p in ax.patches:
    ax.annotate(f'{p.get_width()}', (p.get_width(), p.get_y() +
p.get_height() / 2.), va='center', fontsize=12, color='black', xytext=(5,
0), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Contagem')
plt.ylabel('Status pendente')
plt.title('Contagem de Status do Vencimento por Status pendente')

# Exiba o gráfico
plt.show()

```

```

plt.figure(figsize=(17, 8)) # Ajuste o tamanho conforme necessário

# Count Plot da variável target com barras na horizontal
ax = sns.countplot(data=df_sem_duplicatas, y='Status da Primeira Resposta
Atrasada', hue='Status do vencimento', orient='h')

# Adicione rótulos aos topos das barras
for p in ax.patches:
    ax.annotate(f'{p.get_width()}', (p.get_width(), p.get_y() +
p.get_height() / 2.), va='center', fontsize=12, color='black', xytext=(5,
0), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Contagem')
plt.ylabel('Status da Primeira Resposta Atrasada')
plt.title('Contagem de Status do Vencimento por Status da Primeira Resposta
Atrasada')

# Exiba o gráfico
plt.show()

plt.figure(figsize=(17,4))
# Count Plot da variável target
ax = sns.countplot(data=df_sem_duplicatas, x='Dia_da_semana', hue='Status
do vencimento', orient='v')

# Inclua rotação nos rótulos do eixo x (categorias)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90)

# Adicione rótulos aos topos das barras
for p in ax.patches:
    ax.annotate(f'{p.get_height()}', (p.get_x() + p.get_width() / 2.,
p.get_height()), ha='center', va='center', fontsize=12, color='black',
xytext=(0, 5), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Departamento')
plt.ylabel('Contagem')
plt.title('Contagem de Status do Vencimento por Departamento')

# Exiba o gráfico
plt.show()

```

### 1-Comparativo % em relação a variável target

```

df_group1 = df_sem_duplicatas.groupby(['Categoria', 'Status do
vencimento']).agg({'Identificação da solicitação' :
'nunique'}).reset_index()
df_group1 = df_group1.sort_values(by='Status do vencimento')
df_group1

```

```

df_group2 = df_sem_duplicatas.groupby('Status do
vencimento').agg({'Identificação da solicitação' :
'nunique'}).reset_index()
df_group2

df_group3 = df_group1.merge(df_group2, on = 'Status do vencimento')
df_group3

df_group4 = df_sem_duplicatas.groupby('Categoria').agg({'Identificação da
solicitação' : 'nunique'}).reset_index()
df_group4

df_group5 = df_group1.merge(df_group4, on = 'Categoria')
df_group5

df_group5['Percentual(%)'] = df_group5['Identificação da solicitação_x'] /
df_group5['Identificação da solicitação_y'] * 100
df_group5

plt.figure(figsize=(15, 10))

# Crie o gráfico de barras
chart = sns.barplot(x='Percentual(%)',
                    y='Categoria',
                    data=df_group5,
                    hue='Status do vencimento',
                    orient='h',
                    palette=["red", "green"])

# Adicione rótulos aos topos das barras
for p in chart.patches:
    chart.annotate(f'{p.get_width():.2f}%', (p.get_width(), p.get_y() +
p.get_height() / 2.), ha='center', va='center', fontsize=12, color='black',
xytext=(5, 0), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Percentual(%)')
plt.ylabel('Categoria')
plt.title('Percentual(%) por Categoria e Status do Vencimento')

# Exiba o gráfico
plt.show()

# subcategoria

df['Subcategoria'].value_counts()

df_group6 = df_sem_duplicatas.groupby(['Subcategoria', 'Status do
vencimento']).agg({'Identificação da solicitação' :
'nunique'}).reset_index()
df_group6 = df_group6.sort_values(by='Subcategoria')
df_group6

```

```

df_group2 = df_sem_duplicatas.groupby('Status do
vencimento').agg({'Identificação da solicitação' :
'nunique'}).reset_index()
df_group2

df_group8 = df_group6.merge(df_group2, on = 'Status do vencimento')
df_group8

df_group9 = df_sem_duplicatas.groupby('Subcategoria').agg({'Identificação
da solicitação' : 'nunique'}).reset_index()
df_group9

df_group10 = df_group6.merge(df_group9, on = 'Subcategoria')
df_group10

df_group10.columns

df_group10['Percentual(%)'] = df_group10['Identificação da solicitação_x']
/ df_group10['Identificação da solicitação_y'] * 100
df_group10

plt.figure(figsize=(15, 10))

# Crie o gráfico de barras
chart = sns.barplot(x='Percentual(%)',
                    y='Subcategoria',
                    data=df_group10,
                    hue='Status do vencimento',
                    orient='h',
                    palette=["red", "green"],
                    ci=None, # Desativa a barra de erro para melhorar a
visualização
                    capsizer=11.1, # Ajuste conforme necessário para
aumentar o espaçamento
                    dodge=True) # Separa as barras por hue

# Adicione rótulos aos topos das barras
for p in chart.patches:
    chart.annotate(f'{p.get_width():.2f}%', (p.get_width(), p.get_y() +
p.get_height() / 2.), ha='center', va='center', fontsize=12, color='black',
xytext=(5, 0), textcoords='offset points')

# Adicione rótulos aos eixos
plt.xlabel('Percentual(%)')
plt.ylabel('Subcategoria')
plt.title('Percentual(%) por Subcategoria')

# Exiba o gráfico
plt.show()

# % por Departamento

```

```

df_group12 = df_sem_duplicatas.groupby(['Departamento', 'Status do
vencimento']).agg({'Identificação da solicitação' :
'nunique'}).reset_index()
df_group12 = df_group12.sort_values(by='Departamento')
df_group12

df_group13 = df.groupby('Status do vencimento').agg({'Identificação da
solicitação' : 'nunique'}).reset_index()
df_group13

df_group14 = df_group12.merge(df_group13, on = 'Status do vencimento')
df_group14

df_group15 = df_sem_duplicatas.groupby('Departamento').agg({'Identificação
da solicitação' : 'nunique'}).reset_index()
df_group15

df_group16 = df_group12.merge(df_group13, on = 'Status do vencimento')
df_group16

df_group17 = df_sem_duplicatas.groupby('Departamento').agg({'Identificação
da solicitação' : 'nunique'}).reset_index()
df_group17

df_group18 = df_group12.merge(df_group17, on = 'Departamento')
df_group18

df_group18['Percentual(%)'] = df_group18['Identificação da solicitação_x']
/ df_group18['Identificação da solicitação_y'] * 100
df_group18

plt.figure(figsize=(15, 10))

# Crie o gráfico de barras
chart = sns.barplot(x='Percentual(%)',
                    y='Departamento',
                    data=df_group18,
                    hue='Status do vencimento',
                    orient='h',
                    palette=["red", "green"],
                    ci=None, # Desativa a barra de erro para melhorar a
visualização
                    capsizes=11.1, # Ajuste conforme necessário para
aumentar o espaçamento
                    dodge=True) # Separa as barras por hue

# Adicione rótulos aos topos das barras
for p in chart.patches:
    chart.annotate(f'{p.get_width():.2f}%', (p.get_width(), p.get_y() +
p.get_height() / 2.), ha='center', va='center', fontsize=12, color='black',
xytext=(5, 0), textcoords='offset points')

```

```

# Adicione rótulos aos eixos
plt.xlabel('Percentual(%)')
plt.ylabel('Departamento')
plt.title('Percentual(%) por Departamento')

# Exiba o gráfico
plt.show()

df_group32 = df_sem_duplicatas.groupby(['Dia_da_semana', 'Status do
vencimento']).agg({'Identificação da solicitação' :
'nunique'}).reset_index()
df_group32 = df_group32.sort_values(by='Status do vencimento')

df_group32.info()

df_group32

df_group33 = df_sem_duplicatas.groupby('Status do
vencimento').agg({'Identificação da solicitação' :
'nunique'}).reset_index()
df_group33

df_group34 = df_group32.merge(df_group33, on = 'Status do vencimento')
df_group34

df_group35 = df_sem_duplicatas.groupby('Dia_da_semana').agg({'Identificação
da solicitação' : 'nunique'}).reset_index()
df_group35

df_group36 = df_group32.merge(df_group35, on = 'Dia_da_semana')
df_group36

df_group36['Percentual(%)'] = df_group36['Identificação da solicitação_x']
/ df_group36['Identificação da solicitação_y'] * 100
df_group36

plt.figure(figsize=(15, 10))

# Crie o gráfico de barras
chart = sns.barplot(x='Percentual(%)',
                    y='Dia_da_semana',
                    data=df_group36,
                    hue='Status do vencimento',
                    orient='h',
                    palette=["red", "green"])

# Adicione rótulos aos topos das barras
for p in chart.patches:
    chart.annotate(f'{p.get_width():.2f}%', (p.get_width(), p.get_y() +
p.get_height() / 2.), ha='center', va='center', fontsize=12, color='black',
xytext=(5, 0), textcoords='offset points')

```

```

# Adicione rótulos aos eixos
plt.xlabel('Percentual(%)')
plt.ylabel('Dia_da_semana')
plt.title('Percentual(%) por Dia da semana e Status do Vencimento')

# Exiba o gráfico
plt.show()

categoricas = df_sem_duplicatas[['Categoria','Subcategoria','Criado
por','Técnico','Local','Status do vencimento','Status da solicitação',
'Dia_da_semana']]

solicitacoes_resolvidas_categorias =
df_sem_duplicatas[df_sem_duplicatas['Status da solicitação'] ==
'Resolvido'][['Categoria', 'Subcategoria', 'Criado por', 'Técnico',
'Local', 'Status do vencimento', 'Status da solicitação', 'Dia_da_semana']]

solicitacoes_resolvidas_categorias =
df_sem_duplicatas[df_sem_duplicatas['Status da solicitação'] ==
'Resolvido'][['Categoria', 'Subcategoria', 'Criado por', 'Técnico',
'Local', 'Status do vencimento', 'Status da solicitação','Status da
Primeira Resposta Atrasada', 'Dia_da_semana']]

solicitacoes_resolvidas_categorias

from sklearn.preprocessing import LabelEncoder

# Crie uma instância do LabelEncoder
label_encoder = LabelEncoder()

# Ajuste o LabelEncoder aos valores únicos na coluna "Categoria"
label_encoder.fit(solicitacoes_resolvidas_categorias['Categoria'])

# Transforme os valores na coluna "Categoria" em valores inteiros
solicitacoes_resolvidas_categorias['Categoria'] =
label_encoder.transform(solicitacoes_resolvidas_categorias['Categoria'])

solicitacoes_resolvidas_categorias.Categoria.value_counts()

solicitacoes_resolvidas_categorias

# Ajuste o LabelEncoder aos valores únicos na coluna "Subcategoria"
label_encoder.fit(solicitacoes_resolvidas_categorias['Subcategoria'])

solicitacoes_resolvidas_categorias.Subcategoria.describe()

solicitacoes_resolvidas_categorias.Subcategoria.value_counts()

```

```

# Transforme os valores na coluna "Categoria" em valores inteiros
solicitacoes_resolvidas_categorias['Subcategoria'] =
label_encoder.transform(solicitacoes_resolvidas_categorias['Subcategoria'])

solicitacoes_resolvidas_categorias.Subcategoria.value_counts()

solicitacoes_resolvidas_categorias.rename(columns={'Criado por':
'Criado_por'}, inplace=True)

solicitacoes_resolvidas_categorias

# Ajuste o LabelEncoder aos valores únicos na coluna "Criado por"
label_encoder.fit(solicitacoes_resolvidas_categorias['Criado_por'])

solicitacoes_resolvidas_categorias.Criado_por.describe()

solicitacoes_resolvidas_categorias.Criado_por.value_counts()

# Transforme os valores na coluna "Categoria" em valores inteiros
solicitacoes_resolvidas_categorias['Criado_por'] =
label_encoder.transform(solicitacoes_resolvidas_categorias['Criado_por'])

solicitacoes_resolvidas_categorias.Criado_por.value_counts()

solicitacoes_resolvidas_categorias

solicitacoes_resolvidas_categorias.rename(columns={'Técnico': 'Tecnico'},
inplace=True)

# Ajuste o LabelEncoder aos valores únicos na coluna "Categoria"
label_encoder.fit(solicitacoes_resolvidas_categorias['Tecnico'])

solicitacoes_resolvidas_categorias.Tecnico.describe()

# Transforme os valores na coluna "Categoria" em valores inteiros
solicitacoes_resolvidas_categorias['Tecnico'] =
label_encoder.transform(solicitacoes_resolvidas_categorias['Tecnico'])

solicitacoes_resolvidas_categorias.Tecnico.value_counts()

solicitacoes_resolvidas_categorias

# Ajuste o LabelEncoder aos valores únicos na coluna "Local"
label_encoder.fit(solicitacoes_resolvidas_categorias['Local'])

solicitacoes_resolvidas_categorias.Local.describe()

```

```

solicitacoes_resolvidas_categorias.Local.value_counts()

# Transforme os valores na coluna "Categoria" em valores inteiros
solicitacoes_resolvidas_categorias['Local'] =
label_encoder.transform(solicitacoes_resolvidas_categorias['Local'])

solicitacoes_resolvidas_categorias.Local.value_counts()

solicitacoes_resolvidas_categorias

solicitacoes_resolvidas_categorias.rename(columns={'Status do vencimento':
'Status_do_vencimento'}, inplace=True)

# Ajuste o LabelEncoder aos valores únicos na coluna "Status do vencimento"
label_encoder.fit(solicitacoes_resolvidas_categorias['Status_do_vencimento'
])

solicitacoes_resolvidas_categorias.Status_do_vencimento.describe()

solicitacoes_resolvidas_categorias.Status_do_vencimento.value_counts()

# Transforme os valores na coluna "Categoria" em valores inteiros
solicitacoes_resolvidas_categorias['Status do vencimento'] =
label_encoder.transform(solicitacoes_resolvidas_categorias['Status_do_venci
mento'])

solicitacoes_resolvidas_categorias.Status_do_vencimento.value_counts()

solicitacoes_resolvidas_categorias

solicitacoes_resolvidas_categorias.rename(columns={'Status da Primeira
Resposta Atrasada': 'Status_da_Primeira_Resposta_Atrasada'}, inplace=True)

# Ajuste o LabelEncoder aos valores únicos na coluna
"Status da Primeira Resposta Atrasada"
label_encoder.fit(solicitacoes_resolvidas_categorias['Status_da_Primeira_Re
sposta_Atrasada'])

solicitacoes_resolvidas_categorias.Status_da_Primeira_Resposta_Atrasada.des
cribe()

solicitacoes_resolvidas_categorias.Status_da_Primeira_Resposta_Atrasada.val
ue_counts()

# Transforme os valores na coluna "Categoria" em valores inteiros
solicitacoes_resolvidas_categorias['Status da Primeira Resposta Atrasada']
=
label_encoder.transform(solicitacoes_resolvidas_categorias['Status_da_Prime
ira_Resposta_Atrasada'])

```

```

solicitacoes_resolvidas_categorias

solicitacoes_resolvidas_categorias.drop('Status da solicitação', axis=1,
inplace=True)

solicitacoes_resolvidas_categorias

# Ajuste o LabelEncoder aos valores únicos na coluna "Dia_da_semana"
label_encoder.fit(solicitacoes_resolvidas_categorias['Dia_da_semana'])

solicitacoes_resolvidas_categorias.Dia_da_semana.describe()

solicitacoes_resolvidas_categorias.Dia_da_semana.value_counts()

solicitacoes_resolvidas_categorias['Dia_da_semana'] =
label_encoder.transform(solicitacoes_resolvidas_categorias['Dia_da_semana']
)

solicitacoes_resolvidas_categorias

lista_de_colunas = solicitacoes_resolvidas_categorias.columns.tolist()

print(lista_de_colunas)

# Atributos (features)
atributos = solicitacoes_resolvidas_categorias[['Categoria',
'Subcategoria', 'Criado_por', 'Tecnico', 'Local', 'Status_do_vencimento',
'Status_da_Primeira_Resposta_Atrasada', 'Dia_da_semana']]

# Variável-alvo (target)
target = solicitacoes_resolvidas_categorias['Status_do_vencimento']

# Concatene os atributos e o target em um único DataFrame
dados = pd.concat([atributos, target], axis=1)

# Crie a matriz de correlação
matriz_correlacao = dados.corr()

# Visualize a matriz de correlação
#print(matriz_correlacao)

import seaborn as sns
import matplotlib.pyplot as plt

# Crie a matriz de correlação

```

```

matriz_correlacao = dados.corr()

# Configure o estilo do gráfico
sns.set(style="white")

# Crie um mapa de calor da matriz de correlação
plt.figure(figsize=(10, 8))
sns.heatmap(matriz_correlacao, annot=True, fmt=".2f", cmap='coolwarm',
            linewidths=0.5)
plt.title('Matriz de Correlação')
plt.show()

# verificar multicolinearidadeNo contexto da matriz de correlação que você
gerou para seus atributos e o target, se houver altas correlações entre os
próprios atributos (que não incluem o target), isso pode indicar
multicolinearidade. Além disso, se houver altas correlações entre um
atributo e o target, isso pode indicar uma forte relação entre o atributo e
o resultado que está sendo previsto.

target = target.astype('category')

# Verificar valores ausentes em cada coluna
valores_ausentes = solicitacoes_resolvidas_categorias.isnull().sum()

# Exibir a contagem de valores ausentes
print(valores_ausentes)

# Exibir informações sobre o DataFrame, incluindo valores ausentes
solicitacoes_resolvidas_categorias.info()

# Substitua 'caminho/do/arquivo.txt' pelo caminho e nome do arquivo
desejado
caminho_arquivo_txt = '14012024v1.txt'

# Salvar o DataFrame em formato .txt
solicitacoes_resolvidas_categorias.to_csv(caminho_arquivo_txt, sep='\t',
            index=False)

Separando em treino e teste

from sklearn.model_selection import train_test_split

# Defina as colunas que serão usadas como features (atributos)
colunas_features = ['Categoria', 'Subcategoria', 'Criado_por', 'Tecnico',
                    'Local', 'Status_da_Primeira_Resposta_Atrasada','Dia_da_semana']

# Features (atributos)
X = solicitacoes_resolvidas_categorias[colunas_features]

# Variável-alvo
y = solicitacoes_resolvidas_categorias['Status_do_vencimento']

```

```
# Divisão dos dados em conjuntos de treinamento e teste (80% treinamento, 20% teste)
X_treino, X_teste, y_treino, y_teste = train_test_split(X, y,
test_size=0.2, random_state=42)
```

```
# Exibir as formas dos conjuntos resultantes
print("Forma do conjunto de treinamento (X):", X_treino.shape)
print("Forma do conjunto de teste (X):", X_teste.shape)
print("Forma do conjunto de treinamento (y):", y_treino.shape)
print("Forma do conjunto de teste (y):", y_teste.shape)
```

```
#identificando
```

```
from sklearn.ensemble import RandomForestRegressor
```

```
# Inicialize o modelo de Random Forest
model = RandomForestRegressor()
```

```
# Treine o modelo
model.fit(X, y)
```

```
importances = model.feature_importances_
```

```
feature_importances_df = pd.DataFrame(importances, index=X.columns,
columns=['Importance'])
```

```
feature_importances_df =
feature_importances_df.sort_values(by='Importance', ascending=False)
```

```
print(feature_importances_df)
```

Escolha um modelo de machine learning adequado para o seu problema. A escolha depende do tipo de tarefa (classificação, regressão, etc.) e das características dos seus dados.

```
from sklearn.linear_model import LinearRegression
```

```
from sklearn.metrics import mean_squared_error, r2_score
```

```
modelo = LinearRegression()
```

```
# Treine o modelo usando os dados de treinamento
modelo.fit(X_treino, y_treino)
```

```
from sklearn.linear_model import LogisticRegression
logmodel = LogisticRegression(solver='lbfgs',max_iter=1000)
logmodel.fit(X_treino,y_treino)
```

```

pred01 = logmodel.predict(X_teste)

pred01

from sklearn.metrics import classification_report
print(classification_report(y_teste,pred01))

#Fazer matriz de confusão

#deve dar dois valores zero e dois valores total

from sklearn.metrics import confusion_matrix
conf_mat = confusion_matrix(y_teste, pred01)
print(conf_mat)

#pegar uma linha que ele não atendeu

# inicio do teste para prever tempo

print(solicitacoes_resolvidas_categorias)

# Crie um DataFrame com a série 'Tempo decorrido'
tempo_decorrido_df = pd.DataFrame({'Tempo decorrido': tempo_decorrido})

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_squared_error, r2_score

modelo = LinearRegression()

# Treine o modelo usando os dados de treinamento
modelo.fit(X_treino, y_treino)

X = sm.add_constant(X)

X.head()

modelo = sm.OLS(y,X)

modelo_v1=modelo.fit()

print(modelo_v1.summary())

```