

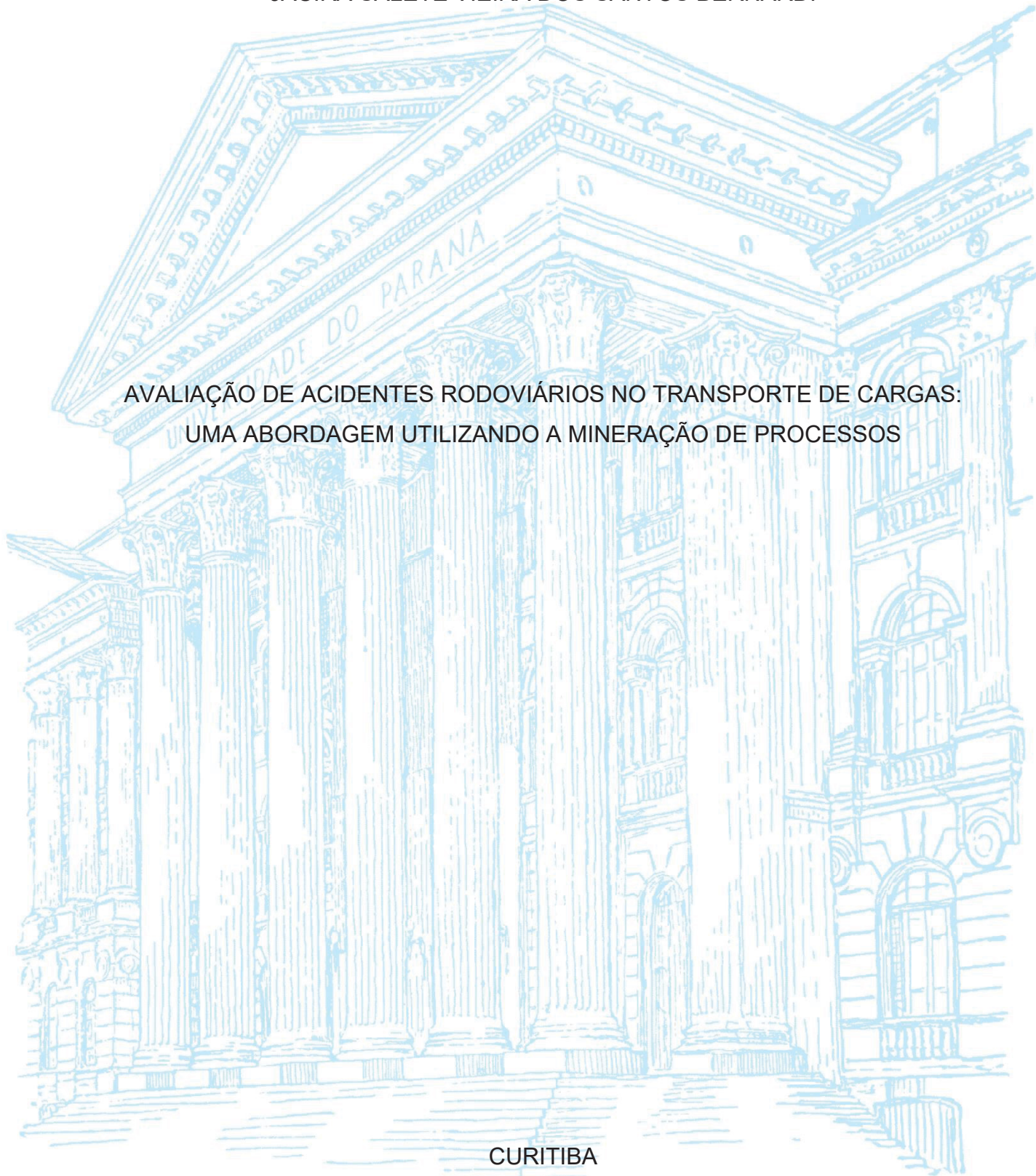
UNIVERSIDADE FEDERAL DO PARANÁ

JACIRA SALETE VIEIRA DOS SANTOS BERNARDI

AVALIAÇÃO DE ACIDENTES RODOVIÁRIOS NO TRANSPORTE DE CARGAS:  
UMA ABORDAGEM UTILIZANDO A MINERAÇÃO DE PROCESSOS

CURITIBA

2024



JACIRA SALETE VIEIRA DOS SANTOS BERNARDI

AVALIAÇÃO DE ACIDENTES RODOVIÁRIOS NO TRANSPORTE DE CARGAS:  
UMA ABORDAGEM UTILIZANDO A MINERAÇÃO DE PROCESSOS

Dissertação apresentada ao curso de Pós-graduação em Gestão de Organizações, Liderança e Decisão, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Gestão de Organizações, Liderança e Decisão.

Orientador: Prof. Dr. Eduardo Alves Portela Santos

CURITIBA

2024

DADOS INTERNACIONAIS DE CATALOCAÇÃO NA PUBLICAÇÃO (CIP)  
UNIVERSIDADE FEDERAL DO PARANÁ  
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS SOCIAIS APLICADAS

Bernardi, Jacira Salete Vieira dos Santos

Avaliação de acidentes rodoviários no transporte de cargas: uma abordagem utilizando a mineração de processos / Jacira Salete Vieira dos Santos Bernardi. – 2024.

1 recurso on-line: PDF.

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão.

Orientador: Eduardo Alves Portela Santos.

1. Mineração de dados. 2. Processos. 3. Transporte de carga. 4. Acidentes de trânsito. 5. Rodovias. I. Santos, Eduardo Alves Portela. II. Universidade Federal do Paraná. Setor de Ciências Sociais Aplicadas. Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão. III. Título.

Bibliotecário Eduardo Silveira – CRB – 9/1921



MINISTÉRIO DA EDUCAÇÃO  
SETOR DE CIÊNCIAS SOCIAIS E APLICADAS  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO DE  
ORGANIZAÇÕES, LIDERANÇA E DECISÃO - 40001016172P9

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GESTÃO DE ORGANIZAÇÕES, LIDERANÇA E DECISÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **JACIRA SALETE VIEIRA DOS SANTOS BERNARDI** intitulada: **AValiação DE ACIDENTES RODOVIÁRIOS NO TRANSPORTE DE CARGAS: UMA ABORDAGEM UTILIZANDO A MINERAÇÃO DE PROCESSOS**, sob orientação do Prof. Dr. EDUARDO ALVES PORTELA SANTOS, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestra está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 29 de Fevereiro de 2024.

Assinatura Eletrônica

01/03/2024 21:40:29.0

EDUARDO ALVES PORTELA SANTOS

Presidente da Banca Examinadora

Assinatura Eletrônica

04/03/2024 15:24:52.0

JOSE MARCELO ALMEIDA PRADO CESTARI

Avaliador Externo (DEPARTAMENTO DE CIÊNCIA E GESTÃO DA INFORMAÇÃO - UFPR)

Assinatura Eletrônica

01/03/2024 20:59:29.0

SILVANA PEREIRA DETRO

Avaliador Externo (40001016)

*Dedico este trabalho aos meus filhos Pedro e Laura Bernardi, que são minha motivação diária. Vocês são minha maior inspiração e me fazem acreditar que posso ir além, mesmo nos momentos mais difíceis. Tudo que faço é pensando no futuro que quero construir para vocês.*

*Espero que, de alguma forma, este trabalho mostre a importância de persistir, de acreditar nos nossos objetivos e seguir em frente, mesmo quando o caminho parece complicado. Vocês me ensinam isso todos os dias, e é por vocês que continuo dando o meu melhor.*

*Com amor e carinho, que este seja só o começo de muitas conquistas, sempre com vocês ao meu lado.*

*Jacira S.V.S. Bernardi*

## AGRADECIMENTOS

Primeiramente, muito obrigada a Deus por me dar a oportunidade de fazer este mestrado, que foi incrível tanto para o meu crescimento pessoal quanto profissional. Agradeço ao meu marido, Everton Bernardi, por me incentivar e apoiar sempre pelo amor e apoio de sempre, aos meus filhos agradeço por entenderem os momentos que precisei estar ausente.

Meu orientador, Eduardo Alves Portela Santos, merece um agradecimento especial pela orientação e sabedoria que foram essenciais para mim.

Aos meus colegas de mestrado, Conrado Pananotti, Rodolfo Bernardi e João Meyer, obrigada por compartilharmos tanto conhecimento. Um obrigado especial para minha amiga Anne Cristine Betoni Cardoso pelo apoio constante, e ao Leandro Tranzola, que mesmo de longe, me inspirou muito.

A todos os professores, especialmente Eduardo Pécora Júnior, Cassius Tadeu Scarpin, Guilherme Francisco Frederico, Simone Cristina Ramos, José, e Egon Walter Wildauer, obrigada pelo profissionalismo e apoio.

Não posso esquecer dos colegas de trabalho, Luciano Farago, Eduardo Alan, meus maiores incentivadores. Leonardo Hatschbach Maciel, pelo suporte técnico e ensinamentos à Bruna Mignon, Tatiane Sousa, Arsenio Flexa, Marcio Molon pelo apoio extra no trabalho.

Um obrigado à minha empresa por toda a flexibilidade e apoio, facilitando essa jornada.

Aos membros da banca de qualificação, Eduardo Pécora Junior e José Marcelo A. P. Cestari e Silvana Detto obrigada pelas contribuições valiosas ao meu projeto.

E, finalmente, obrigada a todos que, de alguma forma, contribuíram para que este trabalho desse certo.

"A educação é a arma mais poderosa que você pode usar para mudar o mundo."

Nelson Mandela

## RESUMO

Este estudo teve como objetivo desenvolver um framework para a aplicação da mineração de processos (MP) na gestão de acidentes rodoviários no transporte de cargas. A partir de uma revisão de literatura, realizada pelo método PRISMA, identificou-se uma lacuna na utilização da MP para analisar dados relacionados a acidentes rodoviários. Para preencher essa lacuna, foi criado o framework PM<sup>2</sup>RCS, utilizando a abordagem metodológica Design Science Research (DSR). O framework foi aplicado em um estudo de caso real para avaliar sua eficácia. A etapa de estruturação da base de dados revelou desafios relacionados à padronização e qualidade das informações. As fontes de dados, oriundas de telemetria, sensores de fadiga e registros de acidentes, apresentavam formatos variados, o que exigiu um extenso trabalho de pré-processamento. Após essa etapa, o framework foi aplicado, resultando na identificação de padrões críticos. O excesso de velocidade e a fadiga foram identificados como os principais fatores de risco, correlacionados a 92% e 38% dos acidentes, respectivamente. A mineração de processos permitiu a visualização de comportamentos de risco que antecedem os acidentes, oferecendo uma base sólida para a implementação de medidas preventivas. Os resultados obtidos demonstraram que o framework PM<sup>2</sup>RCS é uma ferramenta eficaz para a gestão de segurança no transporte rodoviário, proporcionando insights valiosos para a redução de acidentes. Além de possibilitar a análise de grandes volumes de dados, a MP se mostrou útil na identificação de padrões que podem orientar intervenções preventivas. No entanto, o estudo também revelou a necessidade de melhorias contínuas na governança de dados, com especial atenção à padronização e automação dos processos de coleta e análise. Em termos práticos, o framework oferece uma base sólida para a criação de políticas preventivas mais eficazes e orientadas por dados, contribuindo para a segurança operacional no setor de transporte de cargas.

Palavras-chave: Mineração de Processos, Acidentes Rodoviários, Transporte de cargas.

## **ABSTRACT**

This study aimed to develop a framework for the application of process mining (PM) in the management of road traffic accidents in freight transportation. A literature review conducted using the PRISMA method identified a gap in the use of PM for analyzing accident-related data. To address this, the PM2RCS framework was developed using the Design Science Research (DSR) methodology. The framework was applied in a real case study to evaluate its effectiveness. The data structuring phase revealed challenges related to data standardization and quality. Data sources, including telemetry, fatigue sensors, and accident records, had various formats, which required extensive preprocessing work. Afterward, the framework was applied, leading to the identification of critical patterns. Speeding and fatigue were identified as the main risk factors, correlating with 92% and 38% of accidents, respectively. Process mining enabled the visualization of risk behaviors that precede accidents, providing a solid foundation for implementing preventive measures. The results demonstrated that the PM2RCS framework is an effective tool for safety management in road transportation, offering valuable insights into accident reduction. In addition to enabling the analysis of large datasets, PM proved useful in identifying patterns that can guide preventive interventions. However, the study also highlighted the need for continuous improvements in data governance, with special attention to the standardization and automation of data collection and analysis processes. The framework provides a solid foundation for creating more effective, data-driven preventive policies, contributing to operational safety in the freight transportation sector.

**Keywords:** Process Mining, Road Accidents, Freight Transportation.

## LISTA DE FIGURAS

FIGURA 1 – ETAPAS DSC .....	24
FIGURA 2 – OS TRÊS TIPOS BÁSICOS DE MINERAÇÃO DE PROCESSOS .....	28
FIGURA 3 – EXEMPLO LOG DE EVENTOS .....	29
FIGURA 4 – METODOLOGIA PM2.....	30
FIGURA 5 - FASES DA REVISÃO SISTEMÁTICA .....	32
FIGURA 6 - PROCESS MINING PROJECT METHODOLOGY IN ROAD CARGO SAFETY - PM <sup>2</sup> RCS - VERSÃO 2.....	48
FIGURA 7 – PROCESS MINING PROJECT METHODOLOGY IN ROAD CARGO SAFETY PM <sup>2</sup> RCS VERSÃO 3 .....	49
FIGURA 8 – REPRESENTAÇÃO LOGÍSTICA DE LONGO CURSO .....	53
FIGURA 9– ETAPAS DO PREPROCESSAMENTO .....	59
FIGURA 10– FORMATO DADOS SEM INCLUSÃO CASE.....	59
FIGURA 11 – BASE DE DADOS AJUSTADA COM INCLUSÃO DE CASE.....	61
FIGURA 12 – LISTA DE ARQUIVOS DADOS DE EVENTOS TELEMETRIA E FADIGA.....	61
FIGURA 13 – APRESENTAÇÃO FLUXO DOS EVENTOS NO DISCO .....	72
FIGURA 14 – ANÁLISE DE VARIANTES .....	73
FIGURA 15 – ANÁLISE DE VARIANTES .....	74
FIGURA 16 – ANÁLISE ESTATÍSTICA FREQUÊNCIA EVENTOS VIAGENS COM ACIDENTES (Jan a dez/2023).....	76
FIGURA 17 – ESTATÍSTICAS X RECURSO .....	77
FIGURA 18 – ANÁLISE ESTATÍSTICA X TRANSPORTADOR.....	78
FIGURA 19 – ANÁLISE ESTATÍSTICA X TECNOLOGIA .....	79
FIGURA 20 – ANÁLISE ESTATÍSTICA X PLACA.....	80
FIGURA 21 – ANÁLISE COLUNA STATUS.....	81
FIGURA 22 – FLUXO VIAGENS SEM ACIDENTES .....	83
FIGURA 23 – ANÁLISE ESTATÍSTICA FREQUÊNCIA EVENTOS VIAGENS SEM ACIDENTES.....	83
FIGURA 24 – ANÁLISE ESTATÍSTICA FREQUÊNCIA EVENTOS x RECURSO VIAGENS SEM ACIDENTES .....	85
FIGURA 25 – ANÁLISE ESTATÍSTICA FREQUÊNCIA EVENTOS x TECNOLOGIA .....	86

## LISTA DE GRÁFICOS

GRÁFICO 1 – ETAPAS DA SELEÇÃO DOS ESTUDOS .....	36
GRÁFICO 2 – EVOLUÇÃO NA QUANTIDADE DE ACIDENTES (2020 a 2023) .....	65
GRÁFICO 3 – EVOLUÇÃO NA QUANTIDADE DE ACIDENTES MODERADOS e Graves (2020 a 2023) .....	66
GRÁFICO 4 – EVOLUÇÃO NA QUANTIDADE DE ACIDENTES X GRAVIDADE (2020 a 2023).....	67
GRÁFICO 5 – DISTRIBUIÇÃO DO TEMPO DE HABILITAÇÃO DOS MOTORISTAS COM ACIDENTES (2020 a 2023).....	68
GRÁFICO 6 – QUANTIDADE DE ACIDENTES POR ANO E GRAVIDADE RESPONSABILIDADE MOTORISTA (2020 a 2023) .....	69
GRÁFICO 7 – FREQUÊNCIA DAS CAUSAS DE ACIDENTES RESPONSABILIDADE MOTORISTA (2020 a 2023) .....	69
GRÁFICO 8 – MATRIZ DE CONFUSÃO.....	87
GRÁFICO 9 – CURVA ROC PARA AVALIAÇÃO DA ACURACIDADE DO MODELO .....	89

## LISTA DE QUADROS

QUADRO 1 – RESUMO ETAPAS CONSTRUÇÃO DO FRAMEWORK .....	49
QUADRO 2 – DETALHAMENTO FRAMEWORK MINERAÇÃO PROCESSOS GESTÃO DE RISCO DE ACIDENTES RODOVIÁRIOS .....	50
QUADRO 3 – DESCRIÇÃO CAMPOS TABELA DE ACIDENTES.....	64

## LISTA DE TABELAS

TABELA 1 – FONTE DOS DADOS .....	56
TABELA 2 – CLUSTERIZAÇÃO CAUSAS DE ACIDENTES .....	70
TABELA 3 – RESUMO RESULTADOS.....	91

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>16</b>
1.1 PROBLEMA .....	18
1.2 OBJETIVOS .....	19
1.2.1 Objetivo geral .....	19
1.2.2 Objetivos específicos.....	20
1.3 JUSTIFICATIVA .....	20
1.4 ORGANIZAÇÃO DO TRABALHO .....	21
<b>2 METODOLOGIA</b> .....	<b>23</b>
2.1 MÉTODO DA PESQUISA .....	23
2.1.1 Caracterização da pesquisa .....	23
2.1.2 Etapas da Pesquisa .....	24
2.1.2.1 Conscientização .....	24
2.1.2.2 Sugestão .....	25
2.1.2.3 Desenvolvimento .....	25
2.1.2.4 Avaliação .....	25
2.1.2.5 Conclusão.....	25
<b>3 REVISÃO DE LITERATURA</b> .....	<b>27</b>
3.1 FUNDAMENTAÇÃO TEÓRICA.....	27
3.1.1 Mineração de Processos .....	27
3.1.2 METODOLOGIA DE PROJETOS DE MINERAÇÃO DE PROCESSOS (PM <sup>2</sup> )	30
3.2 TRABALHOS RELACIONADOS .....	31
3.2.1 Estratégia de Pesquisa.....	32
3.2.1.1 Critérios de Inclusão.....	34
3.2.1.2 Critérios de Exclusão.....	34
3.2.1.3 Seleção dos estudos .....	35
3.2.1 Discussão.....	44
<b>4 APRESENTAÇÃO DA PROPOSTA</b> .....	<b>47</b>
4.1 PRIMEIRA VERSÃO: IMPLEMENTAÇÃO INICIAL.....	47
4.2 SEGUNDA VERSÃO: APRIMORAMENTO E DETALHAMENTO .....	47
4.3 TERCEIRA VERSÃO: INCORPORAÇÃO DE FEEDBACK E ANÁLISE AVANÇADA.....	48
<b>5 APLICAÇÃO</b> .....	<b>53</b>

5.1 PLANEJAMENTO .....	54
5.1.1 Selecionar processo .....	54
5.1.2 Compor equipe do Projeto.....	54
5.1.3 Definição da questão de pesquisa .....	55
5.1.4 Definição da ferramenta de mineração.....	56
5.2 EXTRAÇÃO.....	56
5.2.1 Identificar origem dos dados .....	56
5.2.2 Extrair os dados .....	57
5.2.3 Transferir conhecimento do processo .....	58
5.3 PRÉ-PROCESSAMENTO .....	58
5.4 PROCESSAMENTO DOS DADOS .....	62
5.5 MINERAÇÃO E ANÁLISE .....	63
5.5.1 Mineração dados Acidentes .....	65
5.5.2 Aplicação da Mineração de Processos.....	71
5.5.2.1 Mineração base de eventos com acidentes.....	73
5.5.2.2 Mineração base de eventos sem acidentes.....	82
5.5.3 Análise de Correlação .....	86
5.6 AVALIAÇÃO .....	90
5.7 MELHORIAS E SUPORTE.....	93
5.7.1 Ações Propostas .....	93
5.7.2 INDICADORES PARA ACOMPANHAMENTO .....	94
5.7.3 MONITORAMENTO CONTÍNUO E SUPORTE OPERACIONAL.....	95
<b>6 APRESENTAÇÃO DOS RESULTADOS .....</b>	<b>96</b>
<b>7 CONCLUSÃO .....</b>	<b>99</b>
<b>REFERÊNCIAS.....</b>	<b>102</b>

## 1 INTRODUÇÃO

De acordo com a OMS (Organização mundial de saúde, 2018) (“World Health Organization, global status report on Road Safety 2018”, 2018), 1,35 milhões de pessoas morrem por ano no mundo em razão de acidentes gerados por veículos motorizados. No Brasil, dados da CNT (“CNT - Confederação Nacional dos Transportes”, 2020) apontam que 63.467 acidentes aconteceram em rodovias federais, 51.865 tiveram vítimas, deixando 71.370 feridos e 5.287 mortos. Além do impacto gerado pela perda das vidas a ocorrência de acidentes gera um elevado prejuízo socioeconômico. De acordo com dados do Conselho Federal de Medicina, o Sistema Único de Saúde gastou quase R\$ 3 Bilhões com acidentes de trânsito no período de 2009 a 2018; outro dado importante refere-se aos prejuízos materiais gerados pelos acidentes envolvendo transporte de cargas, um estudo do Instituto de Pesquisa Econômica Aplicada (IPEA) aponta que em 2017 R\$ 1,7 trilhões foram gastos em sinistros.

Os estudos realizados até o momento sobre acidentes rodoviários buscam identificar os fatores causais, geralmente concentrando-se em demonstrar a influência de fatores específicos, como falhas humanas, condições adversas das estradas ou problemas mecânicos (BAIKEJULI; SHI; HUSSAIN, 2022). No entanto, a maioria das pesquisas tem um enfoque limitado, considerando apenas um fator de cada vez, o que pode subestimar a complexidade das interações entre diferentes variáveis que contribuem para os acidentes (XU, 2019). Uma linha de pesquisa emergente, embora ainda incipiente, envolve o uso de aprendizado de máquina para detectar padrões de risco e propor intervenções preventivas (AHMED; SADULLAH; YAHYA, 2019). Contudo, a literatura mostra uma escassez de estudos que utilizem técnicas de aprendizado de máquina para identificar proativamente as causas dos acidentes e implementar ações preventivas direcionadas para a redução efetiva das ocorrências (KAPLAN et al., 2017).

Nesse contexto, a mineração de processos surge como uma técnica promissora para lidar com a complexidade e a variedade dos dados relacionados aos acidentes rodoviários. Essa abordagem engloba um conjunto de métodos para extrair conhecimento e insights a partir de registros de eventos, permitindo a descoberta de padrões ocultos e a análise detalhada de processos de negócio em execução (VAN DER AALST, 2011). A técnica tem sido amplamente aplicada em diversos setores,

incluindo saúde, manufatura e serviços financeiros, conforme relatado por levantamento realizado pela HSPI Management Consulting (2020), o qual indica a diversidade de aplicações possíveis, embora reconheça que o uso ainda não seja sistemático em todas as áreas.

Estudos sobre acidentes com caminhões pesados indicam que as causas geralmente envolvem a interação de múltiplos fatores. Por exemplo, Baikejuli, Shi e Hussain (2022) concluíram que a maioria dos acidentes fatais com caminhões pesados resulta da ocorrência simultânea de dois ou três fatores contribuintes, como falhas no comportamento do motorista e condições adversas de estrada. De acordo com Xu (2019), a fadiga do motorista é um dos principais fatores que levam a acidentes rodoviários, destacando a importância de monitorar o estado dos motoristas e implementar sistemas de alerta. Vanlaar e Yannis (2006) propuseram um modelo teórico bidimensional que analisa a prevalência e o risco de acidentes, concluindo que, embora o modelo seja promissor, sua aplicabilidade é limitada pela baixa qualidade dos dados disponíveis. Ahmed, Sadullah e Yahya (2019) enfatizaram que a precisão dos dados sobre acidentes é um desafio significativo, com erros médios de 27% na localização dos eventos, 37% nas informações sobre as vítimas, 16% nos dados dos veículos e 19% nas características do ambiente, o que pode comprometer a eficácia das estratégias de prevenção.

Esforços estão sendo intensificados para tornar as estradas mais seguras por meio de tecnologias de assistência ao motorista, que visam prevenir acidentes antes que eles aconteçam. Um estudo realizado por Uma e Eswari (2022, p. 153) avaliou vinte dessas tecnologias, incluindo sistemas que alertam para possíveis colisões, ajustam automaticamente a velocidade e monitoram os pontos cegos dos veículos. Os resultados são promissores: a implementação dessas soluções pode reduzir em até 40% os acidentes envolvendo caminhões pesados, trazendo uma esperança real de segurança para motoristas e passageiros. Em uma abordagem complementar, Kasde e Gugapriya (2016) desenvolveram um protótipo que combina diferentes recursos de segurança, como alertas em tempo real e serviços em nuvem, oferecendo um suporte contínuo aos motoristas.

Por mais avançadas que essas tecnologias sejam, a investigação de acidentes ainda enfrenta desafios. Mesmo com análises detalhadas, o processo de entender o que realmente aconteceu acaba sendo uma tentativa de reconstrução dos eventos, em vez de uma observação direta. Isso significa que, muitas vezes, é

impossível capturar todos os detalhes que contribuíram para o acidente, tornando difícil identificar as causas exatas (Baikajuli; Shi; Hussain, 2022).

Além disso, as abordagens tradicionais de Gerenciamento de Processos de Negócio (BPM) muitas vezes não aproveitam o potencial completo dos dados gerados. Técnicas de Mineração de Dados e Inteligência de Negócios tendem a focar nos números sem considerar o contexto em que os dados são produzidos. A mineração de processos, no entanto, se destaca ao combinar dados e processos, extraindo informações valiosas diretamente dos registros de eventos. Essa técnica é eficaz para entender como os fluxos de trabalho se desenrolam em tempo real, ajudando a identificar padrões importantes e oportunidades de melhorias na segurança (Vergidis; Turner; Tiwari, 2008; Van der Aalst; Weijters, 2004).

## 1.1 PROBLEMA

O aumento significativo no número de acidentes rodoviários, especialmente envolvendo veículos pesados, constitui um problema crítico de segurança pública no Brasil. Estatísticas da Confederação Nacional do Transporte (CNT) evidenciam que a falta de atenção dos condutores figura entre as principais causas desses sinistros, sendo responsável por mais de 25% dos acidentes envolvendo caminhões. Agravando essa situação, a infraestrutura rodoviária deficitária, caracterizada por falhas na sinalização, problemas no pavimento, ausência de acostamento e traçado inadequado, contribui para a elevação do risco de ocorrências.

Neste cenário, as tecnologias de Assistência Avançada ao Condutor (ADAS) e iniciativas de capacitação de motoristas por parte dos embarcadores emergem como esforços significativos para mitigar os riscos associados. No entanto, a implementação eficaz dessas medidas esbarra na gestão e análise dos grandes volumes de dados gerados. A capacidade de interpretar e agir sobre esses dados em tempo real é fundamental para a prevenção de acidentes, mas permanece um desafio notável devido à complexidade e à variedade das informações coletadas.

Um ponto crítico identificado na literatura é a escassez de estudos focados na aplicação da mineração de processos para analisar e utilizar dados no contexto de acidentes rodoviários. Embora essa técnica seja amplamente utilizada em setores como saúde e manufatura, sua aplicação na segurança no trânsito ainda é limitada (DUMAS et al., 2018, p. 45; VAN DER AALST, 2016, p. 23). A mineração de

processos, que permite a extração de conhecimento e insights de logs de eventos, tem se mostrado uma técnica poderosa para revelar padrões ocultos e prever situações de risco, possibilitando intervenções proativas (VERGIDIS; TURNER; TIWARI, 2008, p. 38). No entanto, a aplicabilidade dessa abordagem para melhorar a segurança rodoviária e prevenir acidentes ainda não foi suficientemente explorada, evidenciando uma lacuna na literatura (RODRIGUEZ; LOPEZ, 2022, p. 78; KUMAR; SINGH, 2021, p. 90).

A lacuna na literatura e a necessidade de abordagens inovadoras para a análise de dados no contexto dos acidentes rodoviários destacam o problema central deste estudo: a ausência de um framework estruturado que utilize a mineração de processos para identificar, analisar e agir sobre os fatores de risco associados a acidentes de trânsito. Essa lacuna impede a implementação de medidas preventivas baseadas em dados, limitando significativamente as estratégias de segurança e gestão de riscos.

Diante desse cenário, este estudo propõe o desenvolvimento de um framework baseado em mineração de processos, com o objetivo de facilitar a análise dos dados relacionados a acidentes rodoviários. A intenção é oferecer uma ferramenta que possa auxiliar na geração de insights práticos e contribuir como uma abordagem adicional para orientar medidas preventivas, visando aumentar a segurança nas estradas e reduzir os índices de acidentes.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo geral

O objetivo deste estudo é desenvolver um framework baseado em técnicas de mineração de processos para apoiar na redução de acidentes no transporte rodoviário de cargas. A proposta é integrar a coleta, análise e interpretação de dados de eventos de transporte, com o intuito de identificar padrões e antecipar riscos de acidentes, apoiando a implementação de medidas preventivas e otimizando as respostas em emergências. Além disso, espera-se que o framework ofereça suporte à tomada de decisões fundamentadas em evidências, contribuindo para aumentar a segurança, a eficiência e a sustentabilidade das operações de transporte de cargas.

### 1.2.2 Objetivos específicos

1. Identificar e selecionar as metodologias de mineração de processos mais adequadas para a análise de dados no contexto de acidentes rodoviários, com foco no transporte rodoviário de cargas.
2. Definir uma estrutura de dados otimizada para o pré-processamento das informações geradas pelas tecnologias embarcadas, como sistemas de telemetria e sensores de fadiga, garantindo a qualidade e a integridade dos dados.
3. Aplicar técnicas de mineração de processos para analisar e interpretar os eventos registrados pelos equipamentos de telemetria e sensores de fadiga, identificando padrões e anomalias relacionados a situações de risco.
4. Elaborar e implementar um framework específico para iniciativas de mineração de processos no transporte rodoviário de cargas (TRC), avaliando sua eficácia e propondo melhorias para a gestão de acidentes.

### 1.3 JUSTIFICATIVA

Os acidentes rodoviários envolvendo veículos de carga no Brasil têm gerado preocupações significativas, não apenas pela perda de vidas humanas, mas também pelos elevados impactos sociais e econômicos. De acordo com o Conselho Federal de Medicina (2018), o Sistema Único de Saúde gastou quase R\$ 3 bilhões em atendimentos a vítimas de acidentes de trânsito ao longo de uma década. Além disso, os prejuízos econômicos são substanciais, com o Instituto de Pesquisa Econômica Aplicada (IPEA, 2017) estimando um custo de R\$ 1,7 trilhão em sinistros no setor de transporte de cargas em um único ano. Esses números mostram a gravidade do problema e a urgência em buscar formas mais abrangentes de enfrentar os riscos associados.

Embora existam iniciativas e tecnologias voltadas para a segurança rodoviária, os métodos tradicionais de prevenção têm se mostrado limitados diante da complexidade dos acidentes de trânsito. Esse contexto evidencia a necessidade de

explorar caminhos adicionais e complementares, que possam agregar novas camadas de análise e prevenção às estratégias já adotadas. O desenvolvimento de metodologias que utilizem técnicas de mineração de processos para identificar padrões e antecipar riscos se apresenta como uma alternativa promissora para ampliar as formas de lidar com os desafios da segurança no transporte rodoviário.

A presente pesquisa busca justamente contribuir para essa lacuna, adaptando uma abordagem que integra a análise de dados e a gestão de processos. Ao propor novas formas de explorar os dados gerados por tecnologias embarcadas nos veículos, o estudo visa não apenas fortalecer as práticas preventivas, mas também oferecer soluções mais adaptáveis e eficazes para o contexto do transporte rodoviário de cargas.

Além disso, ao promover o uso de metodologias complementares, esta pesquisa pretende fornecer alternativas acessíveis e práticas para profissionais do setor, auxiliando na superação das limitações enfrentadas pelas abordagens tradicionais. Dessa forma, o estudo não só agrega valor às estratégias já existentes, mas também contribui para a criação de um ambiente de transporte mais seguro e sustentável, com benefícios que se estendem tanto para a segurança pública quanto para a economia.

#### 1.4 ORGANIZAÇÃO DO TRABALHO

Esta pesquisa está estruturada em sete capítulos, visando uma exposição clara e sistemática da pesquisa realizada. A organização do trabalho é detalhada a seguir:

##### Capítulo 1: Introdução

Este capítulo estabelece o contexto da pesquisa, delineando a problemática que motiva o estudo. Objetivos gerais e específicos são definidos para orientar a investigação. A justificativa da pesquisa é apresentada, destacando sua relevância acadêmica e prática.

##### Capítulo 2: Metodologia

Descreve o arcabouço metodológico adotado, incluindo a abordagem de pesquisa, métodos de coleta e análise de dados. Este capítulo também detalha o

planejamento da aplicação da metodologia, esclarecendo como os objetivos serão alcançados.

### Capítulo 3: Referencial Teórico

Apresenta uma revisão abrangente da literatura sobre mineração de processos, estabelecendo os fundamentos teóricos que suportam a pesquisa. Discussões sobre trabalhos correlatos fornecem um panorama das contribuições e lacunas existentes. Conceitos e métodos relevantes para o estudo são explorados em profundidade.

### Capítulo 4: Proposta de Framework

Este capítulo detalha a proposta de um framework para aplicação da mineração de processos no contexto específico de gestão de acidentes no transporte rodoviário de cargas. A estrutura, componentes e funcionalidades do framework são apresentados e justificados.

### Capítulo 5: Estudo de Caso

Ilustra a aplicação do framework proposto através de um estudo de caso, demonstrando sua utilidade e eficácia. A seleção do caso, processo de implementação e análise dos dados são discutidos, oferecendo insights práticos sobre a aplicabilidade do framework.

### Capítulo 6: Resultados

Analisa os resultados obtidos com a aplicação do framework, avaliando seu impacto na gestão de acidentes no transporte rodoviário de cargas. Este capítulo discute as implicações dos achados, sua contribuição para o campo de estudo e sua relevância prática.

### Capítulo 7: Conclusão e Trabalhos Futuros

Conclui o estudo, resumindo os principais achados e suas contribuições. Limitações da pesquisa são reconhecidas e sugestões para trabalhos futuros são propostas, visando a continuidade do avanço na área de mineração de processos aplicada à gestão de acidentes rodoviários.

## 2 METODOLOGIA

Para entendimento de como se chegou à proposta apresentada, nesta seção será descrito o método de pesquisa abordando suas características e método de trabalho indicando as etapas e procedimentos adotados.

### 2.1 MÉTODO DA PESQUISA

A pesquisa foi baseada no método Design Science Research (DSR), amplamente reconhecido como um método eficaz para conduzir pesquisas tecnológicas e científicas que buscam soluções práticas e rigorosas (HEVNER et al., 2004). A DSR é vista como uma metodologia adequada para orientar investigações em áreas como tecnologia, engenharia e sistemas de informação, ao focar na criação e avaliação de artefatos que resolvam problemas identificados no contexto real (PEFFERS et al., 2007).

O método segue etapas claras e estruturadas, que começam com a identificação do problema e prosseguem com o desenvolvimento e avaliação de artefatos, sendo uma abordagem essencialmente iterativa (WIERINGA, 2014). No processo de aplicação do DSR, é necessário garantir que os artefatos projetados possam ser avaliados em termos de sua utilidade e eficácia, utilizando tanto experimentos quanto estudos de caso, dependendo da complexidade do problema investigado (DRESCH; LACERDA; ANTUNES JÚNIOR, 2015).

#### 2.1.1 Caracterização da pesquisa

Natureza “Aplicada”: o objetivo é voltado à geração de conhecimentos aplicados ao contexto da gestão de acidentes rodoviários no transporte de cargas, focados na geração de insights descobertos utilizando dados gerados por tecnologias de assistência ao motorista.

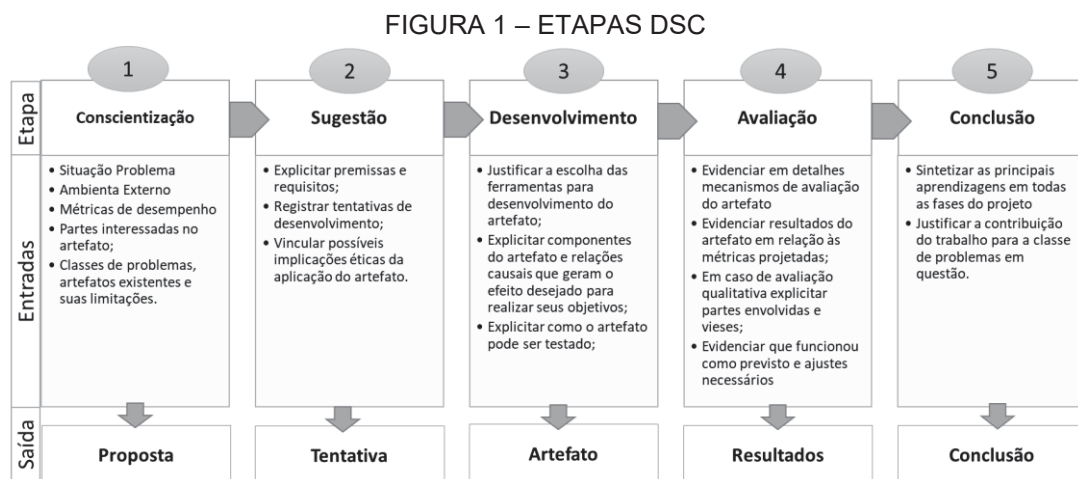
Abordagem mista “Qualitativa” e “Quantitativa”: a pesquisa possui essa classificação mista dado que serão utilizados dados de conhecimento do processo e

estudo dos dados. Por fim, há a generalização da solução para uma certa classe de problemas (CHALMERS, 1993).

A pesquisa também envolve o estudo dos registros de eventos de uma grande multinacional do ramo alimentício de maneira que se permita validar o modelo proposto.

### 2.1.2 Etapas da Pesquisa

A aplicação será baseada no modelo proposto por Hevner, 2007 e Lacerda, 2013 seguindo as etapas de Conscientização, Sugestão, desenvolvimento, avaliação e conclusão conforme descrito na Figura 1.



FONTE: Adaptado de DRESCH et al. (2015).

#### 2.1.2.1 Conscientização

Nesta etapa é realizado o entendimento do problema incluindo a justificativa e relevância do estudo. São identificados os principais problemas existentes, partes interessadas e a definição de métricas/critérios de aceitação. Também são mapeados os estudos já desenvolvidos que utilizam a mineração de processos focada na gestão / redução de acidentes rodoviários de longa distância realizando uma revisão sistemática da literatura baseada no método PRISMA (MOHER, 2010). Adicionalmente foram avaliados estudos sobre causas e tecnologias utilizadas para

redução de acidentes. O resultado esperado é apresentação da proposta do modelo para aplicação. A proposta de artefato.

#### 2.1.2.2 Sugestão

Nesta etapa são definidas as premissas e requisitos do artefato assim como o registro de todas as tentativas de desenvolvimento do artefato. Foram mapeados riscos e implicações da aplicação do artefato.

#### 2.1.2.3 Desenvolvimento

Na etapa de desenvolvimento é realizada a justificativa da escolha das ferramentas para desenvolvimento do artefato. Aqui são descritos os componentes do artefato e as relações causais que geram o efeito desejado para realizar o objetivo. Para avaliar a efetividade do modelo será testada na aplicação em um caso real em uma multinacional de grande porte do ramo alimentício, utilizando os dados de eventos gerados pelas tecnologias embarcadas nos veículos (Rastreador, Telemetria e Sensores de fadiga). São registradas todas as tentativas de aplicação com registro de efetividade e entendimento de novas necessidades de adequação do modelo. A coleta de dados é realizada através das tecnologias de telemetria e fadiga da frota que possui equipamento instalado, a leitura é realizada de maneira automática e recebida através das tecnologias de rastreamento. A análise dos dados é realizada utilizando uma Plataforma de Mineração de Processos.

#### 2.1.2.4 Avaliação

A avaliação da efetividade da utilização do método é realizada com a apresentação dos resultados para os gestores da empresa comparando o modelo tradicional de análise com a aplicação do artefato. Aqui é avaliado o atendimento às métricas de avaliação de desempenho além de entrevistas com usuários.

#### 2.1.2.5 Conclusão

Nesta etapa é realizada a síntese da pesquisa apresentando as descobertas realizadas em cada etapa do estudo. Um dos resultados esperados dessa etapa é a generalização para a aplicação seja possível em outras operações.

### 3 REVISÃO DE LITERATURA

#### 3.1 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica deste estudo apresenta conceitos da Mineração de Processos e a metodologia PM2. Estes conceitos são aplicados para explorar e resolver problemas associados aos acidentes rodoviários no transporte de cargas. Através da mineração de processos, este estudo visa descobrir padrões e modelos que possam identificar riscos potenciais e ineficiências nos processos existentes. Simultaneamente, a metodologia DSR é empregada para desenvolver e validar artefatos que possam mitigar esses riscos, promovendo intervenções eficazes no contexto real do transporte de cargas. Este enfoque não apenas melhora a compreensão dos processos logísticos envolvidos, mas também facilita a implementação de soluções práticas e orientadas a dados para aumentar a segurança e a eficiência no transporte rodoviário de cargas.

##### 3.1.1 Mineração de Processos

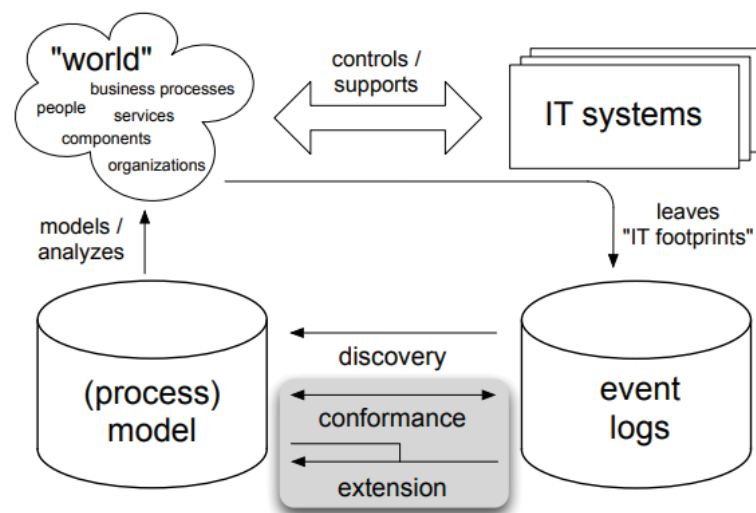
A mineração de processos (MP) pode ser descrita como uma técnica de análise de dados de processos que podem ser usados para produzir modelos formais correspondentes a execuções reais do processo (COOK; WOLF, 1998). Consiste em um conjunto de técnicas que combinam aspectos de modelagem e análise de processos de negócio com mineração de dados e *machine learning* (AILENEI et al., 2011).

As técnicas de mineração de processo estão orientadas para extrair conhecimento a partir de registros de eventos, disponíveis nos sistemas de informação existentes. Estas técnicas oferecem novos meios para descobrir, monitorar e melhorar processos num vasto domínio de aplicações (VOSSEN, 2012). A mineração de processos pode ser dividida em três fases (AALST, 2011): descoberta, conformidade e aprimoramento.

A fase de descoberta analisa o modelo inicial do processo, que pode ser obtido manualmente, usando uma notação para modelagem de processos (AALST et al., 2011) ou através da análise de um log de eventos, minerando todos os possíveis

caminhos de resolução. A fase de conformidade busca medir o nível de alinhamento entre o modelo de descoberta do processo e as informações obtidas no log de eventos e verificar a existência de possíveis desvios. Também nessa etapa, o modelo pode ser estendido, incluindo, por exemplo, tempos e frequências. Assim, é possível detectar problemas, mediante uso de mineração de decisão (ROZINAT; AALST, 2006) e extrair informações de tempo de duração dos processos executados. Já a fase de aprimoramento possibilita utilizar informações adicionais contidas no log de eventos (AALST, 2011), de forma que, ao final do processo, se tenha um modelo integrado de processos que permite ao usuário visualizar múltiplas perspectivas e possibilidades (AALST, 2011). Essas 3 fases: Descoberta, Conformidade e Aprimoramento constituem os três tipos básicos de Mineração de Processos conforme ilustrado na Figura 2.

FIGURA 2 – OS TRÊS TIPOS BÁSICOS DE MINERAÇÃO DE PROCESSOS



Fonte: Aalst et al. (2011)

A Figura 2 ilustra a interação entre o mundo dos processos de negócio, os sistemas de TI que registram os eventos e os logs dos eventos que são usados para criar, verificar e melhorar os modelos com aplicação da MP. Na sequência são detalhados os principais conceitos apresentados na Figura 2.

Van der Aalst and van Hee (2012) descrevem alguns conceitos básicos aplicáveis a todos os processos no escopo da Mineração de Processos:

- Processo de negócio: consiste em um número de tarefas e um conjunto de condições que determina a ordem delas. Um processo também pode ser chamado de procedimento.

- Roteamento: é o caminho ao longo de ramos em particular de um processo, determinando quais tarefas devem ser realizadas, e em qual ordem, considerando que, no geral, pode existir mais de um caminho.

- Tarefa: é uma unidade lógica de trabalho que é realizada por um recurso. Não necessariamente o recurso executa integralmente a tarefa, mas ele é o responsável por ela.

- Case: é o elemento que é produzido ou modificado num trabalho. Também pode ser referenciado como serviço, produto ou item. Cada case tem uma identidade única, além de ter um tempo de vida limitado.

- Recurso: é o nome genérico de uma pessoa, máquina, grupo de pessoas ou máquinas que são designadas tarefas para que as executem.

- Atividade: é a execução de uma tarefa por um recurso.

Na Figura 3 é apresentado um exemplo de log de eventos extraído da base de eventos de telemetria.

FIGURA 3 – EXEMPLO LOG DE EVENTOS

Load	Evento	Início evento	Origem	Acidente	Cavalo Codificado	Motorista Codificado	Tecnologia
2485329	Frenagem Brusca	03/01/2022 04:25	Telemetria	0	CAV1000	MOT3000	Tecnologia 1
2482299	Excesso Velocidade	03/01/2022 04:44	Telemetria	0	CAV1001	MOT3001	Tecnologia 2
2482299	Excesso Velocidade	03/01/2022 04:57	Telemetria	0	CAV1001	MOT3001	Tecnologia 2
2482299	Excesso Velocidade	03/01/2022 05:04	Telemetria	0	CAV1001	MOT3001	Tecnologia 2
2482299	Excesso Velocidade	03/01/2022 05:11	Telemetria	0	CAV1001	MOT3001	Tecnologia 2
2482299	Excesso Velocidade	03/01/2022 05:13	Telemetria	0	CAV1001	MOT3001	Tecnologia 2
2482299	Excesso Velocidade	03/01/2022 05:22	Telemetria	0	CAV1001	MOT3001	Tecnologia 2
2482299	Excesso Velocidade	03/01/2022 05:22	Telemetria	0	CAV1001	MOT3001	Tecnologia 2
2482299	Excesso Velocidade	03/01/2022 05:22	Telemetria	0	CAV1001	MOT3001	Tecnologia 2
2487680	Excesso Velocidade	03/01/2022 05:40	Telemetria	0	CAV1002	MOT3002	Tecnologia 2
2486732	Excesso Velocidade	03/01/2022 05:48	Telemetria	0	CAV1003	MOT3003	Tecnologia 2
2487545	Frenagem Brusca	03/01/2022 06:00	Telemetria	0	CAV1004	MOT3004	Tecnologia 1
2488338	Excesso Velocidade	03/01/2022 06:07	Telemetria	0	CAV1005	MOT3005	Tecnologia 1
2488338	Frenagem Brusca	03/01/2022 06:21	Telemetria	0	CAV1005	MOT3005	Tecnologia 1

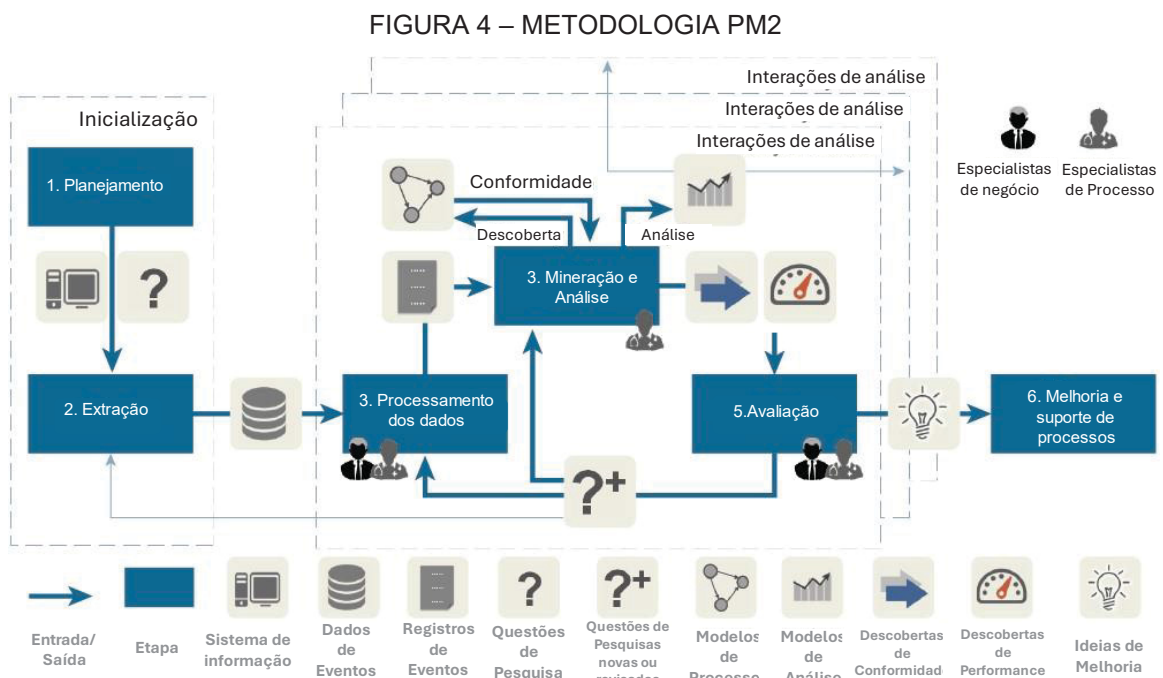
Fonte: O autor (2024)

A Figura 3 apresenta uma tabela que contém colunas essenciais para o acompanhamento e análise dos eventos, como o tipo de evento, seu início, a origem do evento, e dados codificados que identificam tanto o veículo ("cavalo codificado") quanto o motorista responsável. Além disso, inclui informações sobre a tecnologia

utilizada para monitorar esses eventos e qualquer acidente ocorrido durante a operação.

### 3.1.2 METODOLOGIA DE PROJETOS DE MINERAÇÃO DE PROCESSOS (PM<sup>2</sup>)

A metodologia PM2 (Process Mining Project Methodology) é uma abordagem sistemática para conduzir projetos de mineração de processos dentro das organizações. Esta metodologia foi proposta por van der Aalst (2015) em seu trabalho seminal sobre mineração de processos, no qual o autor destaca a importância de um guia estruturado para a implementação eficaz dessa técnica. A PM2 foi desenvolvida para orientar os usuários em todas as fases de um projeto de mineração de processos, desde a concepção até a implementação e o monitoramento de melhorias (VAN DER AALST, 2015), o Framework é apresentado na Figura 4.



FONTE: Traduzido de VAN ECK et al., (2015)

A Figura 4 apresenta uma visão geral da metodologia PM2 para mineração de processos, apresentando um roteiro estratégico para conduzir um projeto desde sua concepção até a implementação das melhorias. A abordagem inicia-se com a etapa de Planejamento (1), que estabelece as bases do projeto, definindo os objetivos e alocando os recursos necessários (VAN DER AALST, 2016). Em seguida, ocorre a

Extração (2), caracterizada pela coleta dos dados essenciais dos sistemas de informação (VAN DER AALST; ADRIANSYAH; MEDEIROS, 2011). Posteriormente, na etapa de Processamento de Dados (3), os dados são limpos e organizados para análise, sendo essa uma fase crucial para garantir a integridade do projeto (ROZINAT; VAN DER AALST, 2008). A fase de Mineração e Análise (4) é onde as técnicas de mineração de processos são aplicadas para extrair modelos descritivos e prescritivos (VAN DER AALST, 2015). Após a mineração, ocorre a Avaliação (5), onde os resultados são analisados meticulosamente para verificar se estão alinhados com os objetivos estabelecidos. Finalmente, a fase de Melhoria e Suporte ao Processo (6) consiste na aplicação das melhorias identificadas e na prestação de suporte contínuo, garantindo que as alterações implementadas sejam sustentáveis e eficazes (VAN DER AALST, 2015). Cada fase da metodologia PM2 é projetada para fornecer saídas específicas que contribuem diretamente para o avanço em direção ao objetivo final do projeto. As entradas e saídas de cada fase são claramente indicadas, oferecendo uma visão sistemática e estruturada para a execução de projetos de mineração de processos.

O diagrama serve como um guia para os praticantes da mineração de processos e é uma representação visual dos conceitos descritos no trabalho de Van Eck et al. (2015), que propõem uma estrutura metodológica para garantir que todas as etapas críticas sejam contempladas e realizadas com êxito

### 3.2 TRABALHOS RELACIONADOS

Uma revisão sistemática é uma revisão que parte de uma pergunta formulada de forma clara, utilizando métodos sistemáticos e explícitos para identificar, selecionar e avaliar criticamente pesquisas relevantes, coletar e analisar dados desses estudos que são incluídos na revisão, esta pesquisa será baseada no método PRISMA (MOHER et al., 2010).

A Figura 5 apresenta um fluxograma utilizado para ilustrar o processo de seleção e avaliação de relatórios ou artigos em uma revisão sistemática da literatura. Este fluxograma é típico do método PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), que é um protocolo reconhecido para conduzir e reportar revisões sistemáticas e meta-análises.

FIGURA 5 - FASES DA REVISÃO SISTEMÁTICA



Fonte: MOHER et al., 2010

O processo inicia com a identificação de relatórios ou artigos através de bancos de dados de pesquisa e outras fontes. Os itens identificados são então filtrados para remover duplicatas. Segue-se a seleção dos relatos rastreáveis, onde são excluídos aqueles que não atendem aos critérios de inclusão estabelecidos. Os artigos remanescentes são avaliados em texto completo quanto à sua elegibilidade. Aqueles que não atendem aos critérios de elegibilidade são excluídos, com as razões para a exclusão devidamente documentadas. Os estudos que passam por esta fase de triagem são então divididos em dois grupos: aqueles que são incluídos na síntese qualitativa e aqueles que são adequados para inclusão na síntese quantitativa, como uma meta-análise.

### 3.2.1 Estratégia de Pesquisa

Para conduzir a revisão sistemática, foram utilizadas as bases de dados Web of Science e Scopus, reconhecidas por sua abrangência em trabalhos científicos de diversas áreas do conhecimento. O período de busca foi delimitado para incluir estudos publicados entre 2012 e agosto de 2023, garantindo que as referências sejam atuais e relevantes para o contexto de investigação.

As buscas foram realizadas utilizando um conjunto de palavras-chave selecionadas para cobrir aspectos importantes do tema, incluindo "Process Mining", "Accidents Prevention", "Active Safety", "Driver Assistance", "Road Accidents", "Large Trucks", "Heavy Vehicles", "Heavy Trucks", "Load Transport" e "Transport". No Quadro 1, são apresentados os termos de busca utilizados, bem como os resultados obtidos em cada uma das bases pesquisadas.

QUADRO 1 – TERMOS E RESULTADOS DAS BUSCAS

Base	Código	Termos de busca	Quantidade
scopus	1	"Process mining" and "risk" or "accidents prevention" or "active safety" or "driver assistance" or "road accidents" and "large trucks" or "trucks" or "heavy vehicles" or "heavy trucks" or "load transport" and "technology"	0
scopus	2	accidents prevention OR "active safety" OR "driver assistance" OR "road accidents" AND "large trucks" OR "trucks" OR "heavy vehicles" OR "heavy trucks" OR "load transport" AND "technology" and not "urban"	137
scopus	3	"Process Mining" and "Transport"	28
scopus	4	"Process Mining" and "Road"	25
web of science	5	Process mining and ("risk" or "accidents prevention" or "active safety" or "driver assistance") and "road accidents" and ("large trucks" or "trucks" or "heavy vehicles")	0
web of science	6	("accidents prevention" OR "active safety" OR "driver assistance" OR "road accidents") AND ("large trucks" OR "trucks" OR "heavy vehicles" OR "heavy trucks" OR "load transport") not urban	133
web of science	7	"Process Mining" and "Transport"	25
web of science	8	"Process Mining" and "Road"	10
<b>Total</b>			<b>358</b>

FONTE: O autor (2023).

Nas bases utilizadas não foram encontrados estudos que utilizem a mineração de Processos com foco na prevenção / redução de acidentes rodoviários. Desta forma

a estratégia de busca foi realizada de forma a contemplar estudos com foco em acidentes rodoviários e artigos que utilizem a mineração de processos no transporte.

#### 3.2.1.1 Critérios de Inclusão

Foram incluídos nesta revisão os estudos que atendem aos seguintes critérios:

- a) Artigos que abordam diretamente as questões de pesquisa, contribuindo para responder aos objetivos propostos;
- b) Estudos que utilizam técnicas de mineração de dados com foco na previsão de acidentes rodoviários;
- c) Trabalhos que aplicam a Mineração de Processos especificamente no contexto do transporte rodoviário.

#### 3.2.1.2 Critérios de Exclusão

Para garantir a relevância e o foco do estudo, os seguintes critérios de exclusão foram aplicados:

- a. Estudos que tratam de acidentes envolvendo pedestres, pois este foco não se alinha ao objetivo principal da pesquisa;
- b. Artigos que abordam acidentes com motociclistas, devido à especificidade das dinâmicas envolvidas nesses casos;
- c. Trabalhos relacionados a acidentes com ciclistas, que apresentam características distintas em comparação com o transporte de cargas;
- d. Mineração de processos aplicada ao transporte marítimo, por se tratar de um contexto diferente do rodoviário;
- e. Estudos que analisam a mineração de processos no transporte ferroviário, devido às particularidades desse setor;
- f. Pesquisas que tratam de acidentes de trânsito de forma genérica, sem uma aplicação específica ao transporte rodoviário de cargas;
- g. Artigos focados no desenvolvimento de novos equipamentos, que fogem ao escopo da análise de processos;

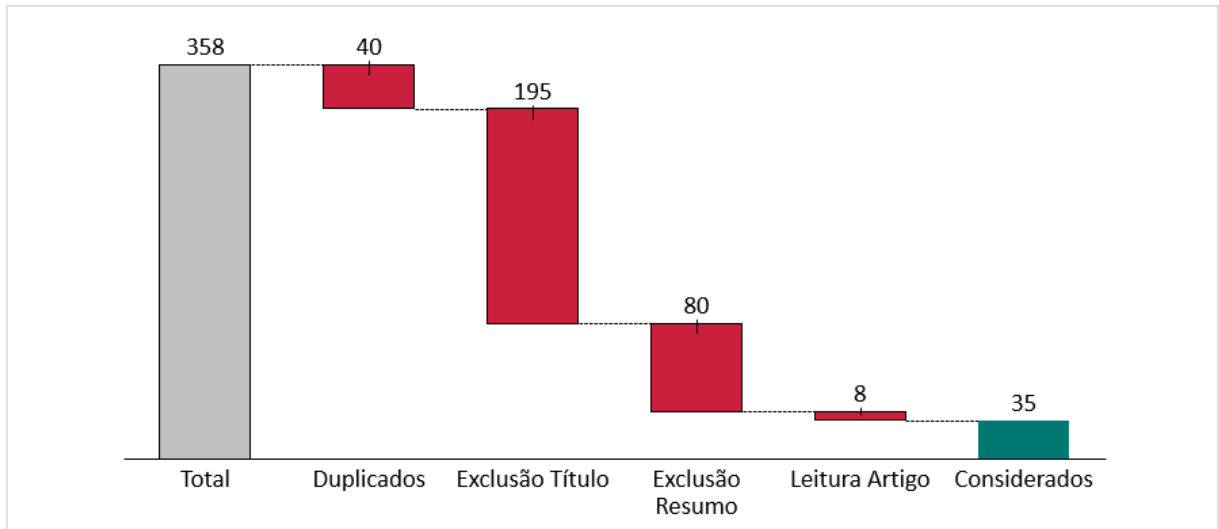
- h. Estudos duplicados, que foram identificados e eliminados para evitar redundância;
- i. Trabalhos que não estavam disponíveis para acesso completo, impedindo sua avaliação aprofundada;
- j. Pesquisas relacionadas à medicina, cujo foco é distinto do objetivo deste estudo.

### 3.2.1.3 Seleção dos estudos

Após a seleção inicial dos candidatos a estudos, os artigos foram analisados com o objetivo de excluir aqueles que não se alinhavam com a pergunta de pesquisa. O processo de seleção seguiu o método PRISMA (MOHER et al., 2009).

- Identificação dos Estudos: Foram realizadas buscas nas bases de dados Web of Science e Scopus em 28 de agosto de 2022, abrangendo o período de 2012 a agosto de 2022. Nesse passo, foram identificados 358 trabalhos relevantes.
- Eliminação de Duplicatas: Os estudos duplicados foram removidos, resultando em 318 trabalhos únicos para análise.
- Triagem pelo Título: Com base na avaliação dos títulos, foram excluídos os artigos que não se relacionavam diretamente ao tema da pesquisa, restando 123 trabalhos.
- Triagem pelo Resumo: Os resumos dos estudos foram analisados para verificar a relevância em relação à pergunta de pesquisa, resultando na exclusão de mais artigos e deixando 43 estudos.
- Análise Final pelo Resumo: Após uma nova revisão dos resumos para garantir a adequação dos estudos à pesquisa, o número final de artigos selecionados foi reduzido para 35 conforme apresentado no Gráfico 1.

GRÁFICO 1 – ETAPAS DA SELEÇÃO DOS ESTUDOS



Fonte: O autor (2023)

O gráfico apresentado na figura 5 ilustra o processo de seleção de estudos para inclusão na revisão sistemática, começando com um total de 358 registros identificados. O processo de exclusão inclui a remoção de 40 registros duplicados, 195 registros após a análise do título, 80 registros após a análise do resumo, e 8 após a leitura completa dos artigos. No final, 35 estudos foram considerados para inclusão na análise final.

A segurança no transporte rodoviário de cargas (TRC) é uma área amplamente estudada, devido à alta incidência de acidentes e seus impactos significativos na economia e na segurança pública. A literatura aborda diferentes métodos para mitigar riscos e otimizar a segurança, variando desde técnicas tradicionais, focadas na análise de acidentes, até abordagens mais modernas, como aprendizado de máquina e tecnologias assistivas. No entanto, a aplicação de técnicas de Mineração de Processos (MP) especificamente para a prevenção de acidentes no TRC ainda é incipiente. Os quadros a seguir sintetizam os principais trabalhos revisados, com base nas abordagens aplicadas e suas limitações. Os artigos foram organizados de acordo com a relevância e o uso de mineração de processos, além de outros critérios específicos relacionados ao tema da pesquisa. O Quadro 2 resume os principais objetivos, contribuições dos artigos selecionados para esta revisão sistemática.

QUADRO 2 – RESUMO ARTIGOS ENCONTRADOS

Referência	Objetivo	Principais Contribuições
Dramski, M. 2017	Propor um modelo para tratar dados faltantes no transporte	Soluções para tratar dados faltantes
Dramski, M. 2018	Criar um modelo confiável de registro de eventos	Diretrizes básicas para registros de eventos
Dorofeev, A.; Kurganov, V.; Fillipova, N.; Pashkova, T. 2020	Garantir a integridade da operação de transporte durante a pandemia	Atuar de maneira remota garantindo níveis de serviço
Porouhan, P. 2020	Avaliar modelos de processo fino dos dados de tráfego rodoviário	Avaliação de falhas no sistema de bilheteria de tráfego
Kurganov, V.; Dorofeev, A.; Gryaznov, M.; Yakimov, M. 2021	Aumentar a confiabilidade do transporte rodoviário de cargas	Ações práticas para melhorar a confiabilidade dos dados
Ribeiro, J.; Tavares, J.; Fontes, T. 2022	Propor uma abordagem da localização dos veículos	Operações logísticas inferidas a partir de dados de geolocalização
Ait-Mlouk, A.; Gharnati, F. 2017	Prever acidentes futuros utilizando mineração de dados	Informações para políticas de prevenção
Gaurav, Alam, Z. 2018	Analisar o tráfego nas rodovias nacionais da Índia	Padrões para dados acidentais e principais causas
Mestri, R.A.; Rathod, R.R. 2020	Remover pontos de concentração de acidentes em rodovias	Medidas para redução de acidentes rodoviários
Garbarino, S.; Traversa, F. 2012	Reduzir riscos de acidentes relacionados à sonolência	Estratégias para reduzir o risco de sonolência
Gregoriades, A.; Mouskos, K.C. 2013	Avaliar o risco de acidentes usando sistemas visuais e dados históricos	Quantificação da segurança no trânsito usando inferência probabilística
De Oña, J.; López, G.; Abellán, J. 2013	Identificar causas dos acidentes rodoviários	Regras de decisão úteis para analistas de segurança viária
Elvik, R. 2013	Avaliar o impacto do uso de medicamentos na ocorrência de acidentes rodoviários	Relação entre acidentes e uso de medicamentos
Barua, D.; Jain, P.; Gupta, J. 2013	Otimizar o projeto de prevenção de acidentes rodoviários	Protótipo de fácil instalação para prevenção de acidentes
Delorme, R.; Lassarre, S. 2014	Avaliar a diferença no risco de fatalidade de acidentes entre França e Grã-Bretanha	Comparações internacionais e fatores discriminatórios
Abdul Shabeer, H.; Abdul Zubar, H. 2014	Propor um dispositivo de segurança para evitar acidentes com uso de celular	Identificação de uso de celular pelo motorista e bloqueio do sinal
Stevenson, M.R.; Elkington, J. 2014	Avaliar fatores que aumentam o risco de acidentes	Identificação de horários e condições associadas ao aumento de risco
Aidman, E.; Chadunow, C. 2015	Avaliar os efeitos do feedback de sonolência em tempo real sobre o desempenho do motorista	Redução da sonolência e maior alerta
Parkkari, K.; Parkkari, I.; Sihvola, N. 2016	Descrever o método de investigação de acidentes na Finlândia	Investigação aprofundada de acidentes no local
Török, Á.; Pauer, G. 2017	Melhorar a segurança rodoviária na Hungria	Oportunidades para melhorar a segurança rodoviária

Yusoff, N.M.; Ahmad, R.F. 2017	Determinar o melhor método para medir distração cognitiva visual	Medição híbrida (físicas e biológicas) é a mais eficaz
Khalil, O.K. 2018	Propor adoção de sistema telemático para segurança viária em Abu Dhabi	Uso da telemática para reduzir mortes e lesões
Nanda, S.; Singh, S. 2018	Identificar causas dos acidentes rodoviários na Índia	Modelo de decisão com múltiplos critérios (AHP difuso)
Mateo, S.S. 2018	Propor um ambiente de fusão de pistas mais seguro	Comprovação de carga mental elevada em fusão de pistas
Olariu, C. 2018	Investigar o uso de cloud-computing para ADAS	Transferência de processamento para a nuvem com 5G
Riaz, F.; Rathore, M.M. 2019	Desenvolver sistema de assistência ao motorista baseado em IoT	Supera ADAS simples sem testes de campo caros
Catalano, M.; Galatioto, F. 2020	Modelar acidentes e apoiar decisão com análise econométrica	Modelos aplicáveis em várias cidades
Chintalapati, S.S. 2020	Avaliar tecnologias IoT para prevenção de acidentes rodoviários	Estruturas baseadas na tecnologia de IoT
Kizito, A.; Semwanga, A.R. 2020	Modelar a ocorrência de acidentes utilizando dados de usuários, policiais e consultores	Modelos qualitativos para diminuir a incidência de acidentes
Shahzad, M. 2020	Analisar acidentes usando SIG (Sistema de informação geográfica)	Técnicas de GIS para simular acidentes e sugerir ferramentas de análise
Karyemsetty, N.; Kumar, K.R. 2020	Desenvolver mecanismo para minimizar acidentes e mitigar efeitos	Precisão de 92,6% para evitar acidentes
Ouazzani-Touhami, K.; Souissi, N. 2020	Avaliar como o monitoramento pode ajudar a prever acidentes	Ciclo de vida de um projeto de simulação
Swamynathan, C.; Palanichamy, M.K.; Jerald, A.R. 2021	Prever comportamento dos fatores humanos na direção	Difusão hibridizada da megatendência com otimização de rede de crenças profundas
Ferreira-Vanegas, C.M.; Velez, J.I.; Garcia-Llinas, G.A. 2022	Identificar principais trabalhos, métodos, tendências e lacunas	Analisou 3888 artigos entre 2000 e 2021
Lu, K.; Karlsson, J.; Dahlman, A.S.; Sjoqvist, B.A.; Candefjord, S. 2022	Investigar a influência da direção automatizada na sonolência do motorista	Monitores de frequência cardíaca vestíveis são úteis para avaliar sonolência

Fonte: O Autor (2023)

Os estudos revisados indicam que a MP tem sido amplamente utilizada para otimizar processos em operações logísticas, como terminais portuários e logística de manufatura, destacando-se as aplicações para descoberta de conformidade e aprimoramento dos fluxos de trabalho (Kedem-Yemini et al., 2018; Rudnitchkaia, 2019). Por exemplo, Kedem-Yemini, Mamon e Mashiah (2018) aplicaram MP para analisar conformidade em terminais portuários, identificando gargalos e melhorando a eficiência operacional. Da mesma forma, Halaška e Šperka (2019) aplicaram a MP na logística de manufatura para automatizar a descoberta de padrões e otimizar os processos produtivos. Esses estudos demonstram o potencial da MP para gerenciar

fluxos operacionais, mas sua aplicação direta para mitigar riscos no TRC ainda é rara, evidenciando uma lacuna na literatura.

Embora a MP seja eficaz para a otimização de processos logísticos, o uso da técnica para a prevenção de acidentes em segurança viária permanece limitado. A literatura aponta que as aplicações de MP para a identificação de padrões críticos de risco em segurança viária ainda são pouco exploradas, o que representa uma oportunidade para a inovação e o desenvolvimento de novos frameworks, como o proposto neste estudo. O Quadro 3 apresenta uma síntese dos artigos que aplicam a MP na logística e a metodologia utilizada em cada estudo.

QUADRO 3: RESUMO DOS ARTIGOS MP NA LOGÍSTICA

<b>Artigo/Autor/ano</b>	<b>Metodologia</b>	<b>Ferramenta</b>	<b>Aplicação</b>
Kedem-Yemini, Mamon & Mashiah (2018)	1 - PM2 2 - Planejamento 3 - Extração 4 - Processamento dos dados 5 - Mineração e análise 6 - Avaliação 7 - Melhoria e suporte	Disco / PROM	Terminais portuários
Rudnitckaia (2019)	1 - Entendimento cenário 2 - Preparação Log de eventos 3 - Pré-processamento 4 - Descoberta do modelo 5 - Análise de oportunidades de otimização 6 - Enriquecimento do modelo	Minerador Fuzzy - PROM	Terminais portuários
Halaška & Šperka (2019)	1 – Aquisição dos logs pelo anylogic 2 – Descrição dos processos de negócio capturados pelo modelo 3 – Técnicas automatizadas de descoberta	NA	Logística de Manufatura
Wang, Hulstijn & Tan (2018)	1 – Mapeamento de informações 2 – Descoberta 3 – Check de conformidade	Fuzzy Miner no ProM	Fiscalização Aduaneira

	4–Análise de desvios/adequação		
Kim (2018)	1 – Definição do problema 2 – Preparação dos dados 3 – Análise do processo 4 - Conclusão	PROM	NA

FONTE: O AUTOR (2023)

A revisão também incluiu estudos focados em técnicas preditivas para melhorar a segurança viária. Trabalhos como os de Ait-Mlouk e Gharnati (2017) utilizaram mineração de dados para prever acidentes com base em padrões históricos, oferecendo subsídios para intervenções preventivas em áreas de maior risco. Mestri e Rathod (2020) exploraram a identificação de pontos críticos de acidentes em rodovias, sugerindo medidas para reduzir a ocorrência de incidentes nesses locais. Esses estudos demonstram o potencial das técnicas preditivas para melhorar a segurança no trânsito, mas não consideram a integração dessas metodologias com a MP para uma análise abrangente dos processos que levam aos acidentes.

Além disso, tecnologias assistivas, como sensores de proximidade e sistemas de alerta para motoristas, têm sido objeto de estudo para a prevenção de acidentes envolvendo veículos pesados. Uma e Eswari (2022) sugerem que essas tecnologias podem reduzir os acidentes em até 40% em cenários específicos, mas ainda enfrentam limitações, como a qualidade inconsistente dos dados e a falta de integração com a análise de processos. Essa lacuna na integração de MP e técnicas preditivas com tecnologias assistivas representa uma oportunidade para desenvolver abordagens mais robustas e orientadas por dados para a segurança no TRC.

A revisão revela que a qualidade dos dados é um desafio constante na aplicação de MP, especialmente em cenários que dependem de registros operacionais e telemáticos. Estudos como os de Dramski (2018) sugerem que a falta de padronização nos registros de dados e a presença de dados ausentes ou imprecisos podem comprometer a eficácia das análises de MP. Para superar essas limitações, é necessário implementar processos rigorosos de verificação e padronização dos dados, além de explorar a utilização de tecnologias emergentes, como a Internet das Coisas (IoT) e sistemas telemáticos para fornecer informações em tempo real.

Apesar dos desafios, a literatura sugere que a combinação de MP com tecnologias emergentes pode fornecer uma abordagem mais holística e proativa para a segurança no TRC. Khalil (2018) e Riaz & Rathore (2019) destacam o potencial de integrar a MP com IoT para capturar a sequência de eventos em tempo real e antecipar intervenções antes que os acidentes ocorram. Essa integração ainda é pouco explorada na literatura, mas representa uma oportunidade significativa para transformar a segurança viária em práticas mais preditivas e baseadas em dados.

O Quadro 4 apresenta uma análise comparativa entre os estudos que aplicam técnicas de mineração de processos em contextos correlatos ao Transporte Rodoviário de Cargas (TRC), como terminais portuários, logística de manufatura e fiscalização aduaneira. Cada estudo examinado segue uma abordagem metodológica específica, variando desde frameworks tradicionais, como o PM<sup>2</sup>, até etapas personalizadas adaptadas às necessidades específicas de cada ambiente operacional

QUADRO 4 – COMPARATIVO ESTAPAS DA METODOLOGIA X AUTOR

PM2 - VAN ECK et al., (2015)	KEDEM-YEMINI; MAMON; MASHIAH (2018)	RUDNITCKAIA (2019)	HALAŠKA; ŠPERKA (2019)	WANG; HULSTIJN; TAN (2018)	Kim (2018)
<b>1. Planejamento</b>	Planejamento	Entendimento do cenário	Não especificado	Mapeamento de informações	Definição do problema
<b>2. Extração de Dados</b>	Extração	Preparação Log de eventos	Aquisição dos logs pelo AnyLogic	Descoberta	Preparação dos dados
<b>3. Processamento de Dados</b>	Processamento dos dados	Pré-processamento	Descrição dos processos de negócio	Check de conformidade	Análise do processo
<b>4. Mineração e Análise</b>	Mineração e análise	Descoberta do modelo	Técnicas automatizadas de descoberta	Análise de desvios/adequação	Conclusão
<b>5. Avaliação</b>	Avaliação	Análise de oportunidades de otimização	Não especificado	Não especificado	Não especificado
<b>6. Melhoria e Suporte</b>	Melhoria e suporte	Enriquecimento do modelo	Não especificado	Não especificado	Não especificado

Fonte: O Autor (2023)

O Quadro 4 compara as etapas do framework PM<sup>2</sup> de van Eck et al. (2015) com outros frameworks e abordagens de mineração de processos propostas por diversos autores. As colunas representam as etapas de diferentes metodologias, incluindo aquelas desenvolvidas por Kedem-Yemini, Mamon, & Mashiah (2018), Rudnitckaia (2019), Halaška & Šperka (2019), Wang, Hulstijn, & Tan (2018), e Kim (2018). O quadro destaca as semelhanças e diferenças na abordagem de cada autor em relação a etapas como planejamento, extração de dados, processamento de dados, mineração e análise, avaliação e melhoria.

A análise revela que quatro dos cinco estudos (80%) incluem uma fase inicial de planejamento ou entendimento do contexto do processo. Apenas Halaška e Šperka (2019) não especificam detalhadamente essa etapa. Todos os estudos (100%) contemplam uma fase de extração ou coleta de dados, sendo que Rudnitckaia (2019) e Halaška & Šperka (2019) enfatizam a preparação e pré-processamento dos dados com maior detalhe. Quatro dos cinco estudos (80%) mencionam o processamento de dados, com Rudnitckaia (2019) abordando explicitamente o pré-processamento. Wang, Hulstijn e Tan (2018) concentram-se mais na conformidade dos dados.

Todos os estudos analisados (100%) incluem uma etapa de análise de mineração de processos, embora a especificidade das técnicas varie. Halaška & Šperka (2019) utilizam técnicas automatizadas, enquanto Kim (2018) apresenta uma abordagem mais geral. No entanto, apenas dois dos cinco estudos (40%) – Kedem-Yemini, Mamon e Mashiah (2018) e Rudnitckaia (2019) – incluem uma fase clara de avaliação dos resultados. Os demais estudos não especificam essa etapa. Da mesma forma, apenas dois estudos (40%) mencionam explicitamente melhorias ou suporte contínuo, sendo novamente Kedem-Yemini, Mamon e Mashiah (2018) e Rudnitckaia (2019) os que abordam essas etapas.

A análise quantitativa destaca que, embora todos os estudos tratem de extração de dados e análise de processos, apenas 80% incluem uma fase de planejamento e somente 40% abordam as etapas de avaliação e melhoria contínua. Isso sugere que a maioria dos estudos não segue uma metodologia tão estruturada e abrangente quanto a PM<sup>2</sup>. A ausência de etapas detalhadas de "Avaliação" e "Melhoria e Suporte" em 60% dos estudos indica uma oportunidade para uma aplicação mais completa e integrada da PM<sup>2</sup> no contexto de segurança no transporte rodoviário de cargas.

A revisão revelou que a mineração de processos ainda é uma metodologia pouco explorada no contexto específico do transporte rodoviário de cargas (TRC). Dos artigos selecionados, cinco abordam transporte e mineração de processos, incluindo os trabalhos de Dramski (2017, 2018), Dorofeev et al. (2020), Porouhan (2020), Kurganov et al. (2021), e Ribeiro et al. (2022). Esses estudos destacam a aplicação da mineração de processos para criar modelos confiáveis de registros de eventos e tratar dados faltantes no setor de transporte e logística. Entretanto, não foram encontrados artigos que utilizassem explicitamente a mineração de processos para estudar acidentes rodoviários, evidenciando uma lacuna significativa na literatura (0 de 35 artigos, 0%).

Em relação à análise de dados e técnicas de predição, 7 dos 35 artigos (20%) abordam essas questões. Autores como Ait-Mlouk e Gharnati (2017), Gaurav e Alam (2018), Mestri e Rathod (2020), Catalano e Galatioto (2020), Shahzad (2020), Swamynathan et al. (2021), e Kizito e Semwanga (2020) demonstram como a mineração de dados pode prever acidentes futuros, fornecendo uma base sólida para o desenvolvimento de políticas de prevenção. Esses estudos apresentaram métodos eficazes para identificar e mitigar pontos críticos de acidentes, reforçando a importância de abordagens baseadas em dados para melhorar a segurança viária.

Estudos de revisão sistemática e meta-análise foram realizados em 2 dos 35 artigos (6%), como os de Elvik (2013) e Ferreira-Vanegas et al. (2022). Esses estudos sintetizam as evidências existentes e identificam lacunas nas pesquisas sobre acidentes rodoviários, oferecendo uma visão abrangente das tendências da literatura e ressaltando a necessidade de mais investigações sobre fatores de risco, como o uso de medicamentos e suas implicações na segurança no trânsito.

A adoção de inovações tecnológicas foi discutida em 6 dos 35 artigos (17%). Trabalhos de Barua et al. (2013), Khalil (2018), Riaz e Rathore (2019), Olariu (2018), Chintalapati (2020), e Karyemsetty e Kumar (2020) exploram a implementação de sistemas telemáticos, dispositivos de segurança e o uso da Internet das Coisas (IoT) e computação em nuvem para desenvolver sistemas avançados de assistência ao motorista (ADAS). Essas tecnologias oferecem soluções eficazes para monitorar e melhorar o comportamento dos motoristas, contribuindo para a prevenção de acidentes.

Estudos focados no comportamento dos motoristas e segurança no trânsito foram abordados em 4 dos 35 artigos (11%). Trabalhos de Garbarino e Traversa

(2012), Stevenson e Elkington (2014), Aidman e Chadunow (2015), e Mateo (2018) enfatizam a importância de estratégias para reduzir a sonolência e melhorar o estado de alerta dos motoristas, além de propor ambientes de fusão de pistas mais seguros, contribuindo para a redução de acidentes.

### 3.2.1 Discussão

A revisão da literatura revela uma lacuna significativa na aplicação de Mineração de Processos (MP) diretamente voltada para a prevenção de acidentes no Transporte Rodoviário de Cargas (TRC). Embora a MP tenha sido amplamente aplicada para otimizar processos logísticos e melhorar a qualidade dos dados, como demonstrado nos estudos de Dramski (2017, 2018) e Dorofeev et al. (2020), sua aplicação no contexto de segurança rodoviária ainda é incipiente. A maioria dos estudos revisados focou em aspectos logísticos, como a continuidade operacional e a eficiência do transporte, sem explorar plenamente o potencial da MP para identificar padrões de risco que possam levar a acidentes.

Os estudos de prevenção de acidentes revisados, como os de Ait-Mlouk e Gharnati (2017), Mestri e Rathod (2020), e Baikajuli et al. (2022), demonstram que técnicas de mineração de dados são eficazes para prever incidentes e identificar pontos críticos em rodovias. No entanto, a falta de integração dessas técnicas com MP representa uma oportunidade não explorada. A MP, quando aplicada de maneira integrada com análise preditiva, poderia fornecer uma visão mais holística dos fatores que levam a acidentes, capturando a sequência exata de eventos e permitindo intervenções mais precisas (Van der Aalst, 2011; Van der Aalst & Weijters, 2004). Ait-Mlouk e Gharnati (2017) mostram que a combinação de técnicas de machine learning com MP pode melhorar significativamente a previsão de eventos críticos, o que sugere que a MP pode desempenhar um papel central na identificação e mitigação de riscos em processos complexos como o TRC. Por exemplo, enquanto os modelos preditivos de mineração de dados identificam áreas de risco, a MP pode ser usada para analisar os processos operacionais subjacentes que contribuem para esses riscos. Isso poderia levar ao desenvolvimento de medidas preventivas mais eficazes, baseadas não apenas na identificação de áreas problemáticas, mas também na compreensão dos processos que as originam.

Apesar das oportunidades, a aplicação da MP para a prevenção de acidentes no TRC enfrenta vários desafios. Primeiro, a qualidade e a consistência dos dados são fundamentais para o sucesso da MP. Estudos como os de Ahmed et al. (2019) apontam erros significativos nos registros de acidentes, o que pode comprometer a eficácia da análise de MP. Garantir a integridade dos dados e superar as barreiras técnicas associadas à sua coleta e processamento é essencial para que a MP possa ser aplicada de maneira eficaz nesse contexto. Além disso, a complexidade dos processos de transporte e a variabilidade nas condições operacionais exigem uma adaptação contínua das técnicas de MP. A metodologia PM2, conforme descrito por Van Eck et al. (2015), oferece uma estrutura iterativa que pode ajudar a mitigar esses desafios, mas sua aplicação requer uma colaboração estreita entre especialistas em MP, profissionais de segurança viária e engenheiros de dados. A integração dessas disciplinas é crucial para desenvolver soluções que não apenas sejam tecnicamente robustas, mas também relevantes e aplicáveis no campo.

A literatura também destaca o papel crescente das tecnologias emergentes, como IoT e sistemas telemáticos, na segurança viária (Khalil, 2018; Riaz e Rathore, 2019). Essas tecnologias fornecem dados em tempo real que podem ser cruciais para a prevenção de acidentes. No entanto, a verdadeira inovação surge quando esses dados são analisados em conjunto com técnicas de MP, permitindo uma análise mais profunda e abrangente dos eventos que levam a acidentes. A combinação de MP com dados de IoT, por exemplo, pode criar sistemas que não só detectam padrões de risco em tempo real, mas também fornecem feedback imediato para motoristas e gestores de frotas. Isso transforma a segurança no TRC de uma abordagem reativa para uma abordagem proativa, onde os riscos são identificados e mitigados antes que os acidentes ocorram.

A lacuna identificada na aplicação de MP para a segurança viária sugere várias direções para futuras pesquisas. Primeiramente, há uma necessidade de estudos que explorem a integração de MP com técnicas de mineração de dados e tecnologias emergentes, como IoT, em projetos de segurança no TRC. Isso inclui o desenvolvimento de frameworks específicos que direcionem de maneira mais específica a aplicação para o contexto de segurança viária, facilitando a aplicação. Além disso, futuras pesquisas devem focar na melhoria da qualidade dos dados utilizados para MP. Isso inclui a padronização dos registros de eventos e o desenvolvimento de técnicas avançadas de limpeza e verificação de dados. A

qualidade dos dados é o alicerce sobre o qual a MP pode ser aplicada de forma eficaz, e sem isso, qualquer análise estará em risco de fornecer resultados imprecisos ou incoerentes.

## 4 APRESENTAÇÃO DA PROPOSTA

O framework PM2-RCS (Process Mining Project Methodology in Road Cargo Safety) foi desenvolvido com o objetivo de aplicar a Mineração de Processos na análise de grandes volumes de dados históricos de acidentes rodoviários no transporte de cargas. Esse framework visa identificar padrões e oferecer recomendações para a melhoria da segurança no transporte rodoviário. Ao contrário das abordagens convencionais, que se concentram em análises isoladas de causas de acidentes, o PM2-RCS utiliza uma análise integrada e preditiva baseada em dados históricos para fornecer insights preventivos.

A seguir, será detalhada a estrutura e o desenvolvimento do framework em suas três versões, além da apresentação da metodologia PM2 e as suas sete etapas principais.

### 4.1 PRIMEIRA VERSÃO: IMPLEMENTAÇÃO INICIAL

Na fase inicial de desenvolvimento do PM2-RCS, o foco principal foi estabelecer uma base sólida de dados e implementar o pré-processamento necessário para a Mineração de Processos. A coleta de dados incluiu informações históricas sobre acidentes, abrangendo informações como condições das estradas, comportamento dos motoristas e detalhes sobre os veículos. Para guiar essa etapa utilizou-se o PM2 conforme Figura 4.

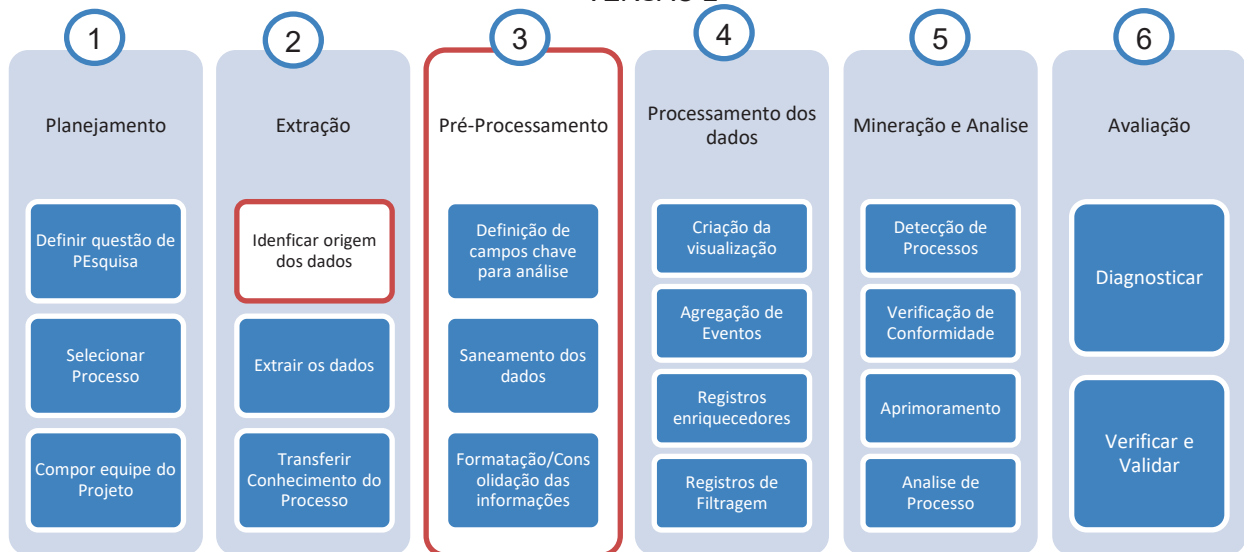
Para garantir que os dados estivessem prontos para análise, foi implementado um processo de pré-processamento, incluindo a remoção de duplicatas, a normalização dos dados e a correção de entradas incompletas ou inconsistentes. O objetivo de preparar os dados para serem processados pelos algoritmos de Mineração de Processos, garantindo que a qualidade da análise não seja comprometida (VAN DER AALST; WEIJTERS, 2004).

Os principais desafios enfrentados nesta fase envolvem a diversidade e qualidade dos dados disponíveis, que exigiram um esforço considerável de limpeza e organização.

### 4.2 SEGUNDA VERSÃO: APRIMORAMENTO E DETALHAMENTO

Na segunda fase, o framework foi expandido com um pré-processamento aprimorado, concentrando-se no detalhamento das atividades necessárias para refinar os dados para a mineração. Nessa fase foram incorporados algoritmos para manipulação e tratamento da base de dados. Esse aprimoramento facilitou uma análise mais profunda, permitindo uma preparação de dados mais eficaz para a identificação de padrões e insights relevantes conforme detalhamento Figura 6.

FIGURA 6 - PROCESS MINING PROJECT METHODOLOGY IN ROAD CARGO SAFETY - PM<sup>2</sup>RCS - VERSÃO 2



FONTE: O Autor (2023).

#### 4.3 TERCEIRA VERSÃO: INCORPORAÇÃO DE FEEDBACK E ANÁLISE AVANÇADA

A terceira versão do framework representou uma etapa de consolidação e expansão significativa. Nesta fase, foram incorporadas as sugestões recebidas da banca avaliadora e dos stakeholders da empresa, levando à inclusão de análises de correlação e algoritmos de predição, durante a etapa de qualificação foi sugerida a inclusão de uma análise de regressão assim como execução de análise preditiva. Além disso, uma análise detalhada da base de dados de acidentes foi realizada, permitindo um entendimento mais profundo das causas e das tendências evolutivas dos acidentes. Essa versão também se beneficiou da integração de feedback específico, resultando em ajustes direcionados que melhoraram a eficácia do framework na prevenção de acidentes. A aplicação de técnicas avançadas de análise

de dados, juntamente com uma abordagem sistemática para revisar e integrar feedback relevante, culminou na formulação do PM2-RCS como uma metodologia adaptada para abordar as questões de segurança no transporte de cargas nesta etapa também foi criado um quadro com detalhamento de cada etapa a fim de facilitar a condução do estudo. No Quadro 4 é apresentado um resumo da agregação realizada em cada fase.

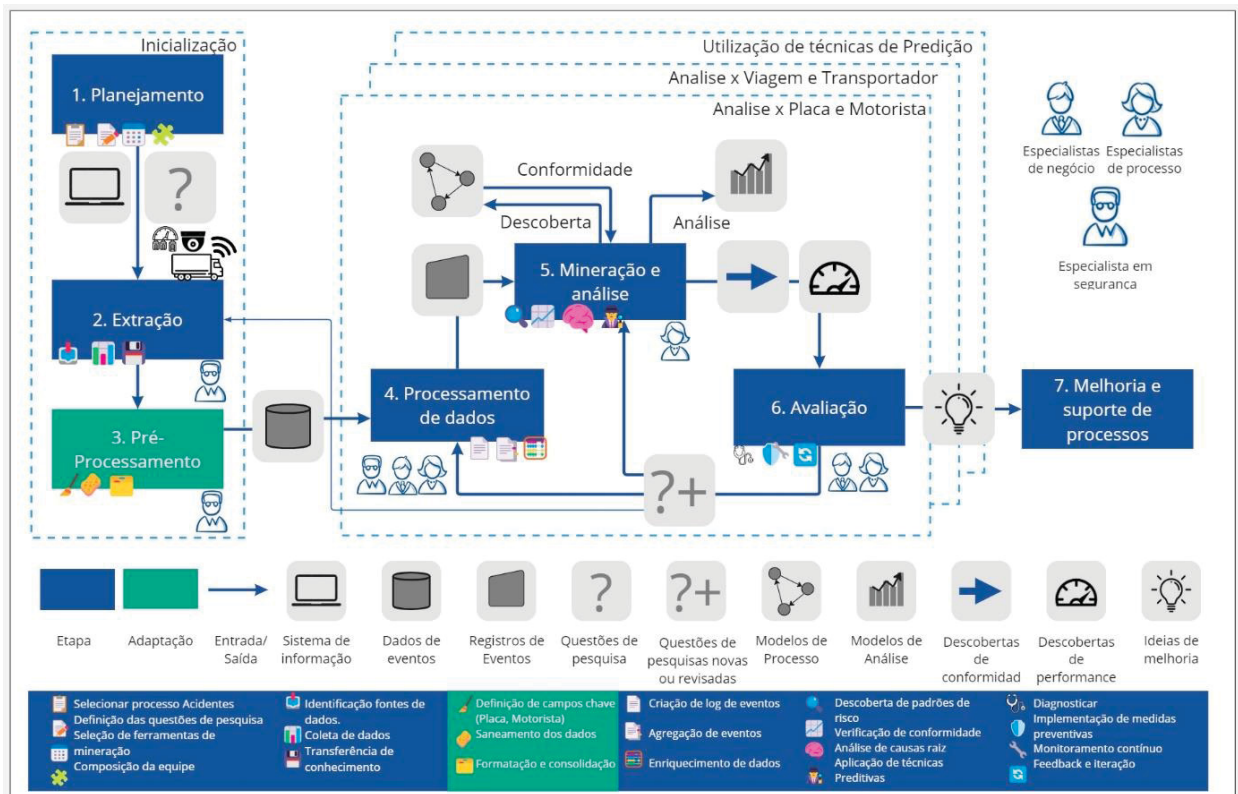
QUADRO 1 – RESUMO ETAPAS CONSTRUÇÃO DO FRAMEWORK

Versão	Descrição	Objetivo / Contribuições
1	Implementação inicial da PM2; identificação de desafios.	Introdução do pré-processamento para adequação de dados.
2	Aprimoramento com pré-processamento detalhado.	Detalhamento de atividades para refino dos dados.
3	Feedback integrado; análise avançada.	Incorporação de análises de correlação, algoritmos de predição, e análise detalhada da base de acidentes.

Fonte: O Autor (2024)

Na sequência é apresentado o resultado do framework com a incorporação das etapas. Conforme detalhado na Figura 7.

FIGURA 7 – PROCESS MINING PROJECT METHODOLOGY IN ROAD CARGO SAFETY PM<sup>2</sup>RCS VERSÃO 3



FONTE: O Autor (2023).

A Figura 7 ilustra o framework PM2-RCS, que integra as etapas de planejamento, extração, pré-processamento, processamento de dados, mineração e análise, avaliação e melhoria e suporte de processos. O diagrama detalha a aplicação de técnicas de predição e conformidade, destacando a interação entre especialistas de negócio, especialistas de processo, e especialistas em segurança. A metodologia PM2-RCS é adaptada para abordar a segurança no transporte rodoviário de cargas, utilizando dados de viagens, motoristas, e eventos para identificar padrões de risco e implementar melhorias contínuas.

O Quadro 5 apresenta o detalhamento de cada uma das etapas e as atividades indicadas, foi idealizado para facilitar a aplicação incluindo os aprendizados durante a etapa de aplicação do estudo de caso.

QUADRO 2 – DETALHAMENTO FRAMEWORK MINERAÇÃO PROCESSOS GESTÃO DE RISCO DE ACIDENTES RODOVIÁRIOS

Fase	Etapa	Descrição das atividades
Planejamento	Selecionar Processo	Definir o processo que será investigado, delimitar claramente onde começa e termina a análise. Serão avaliados todos os eventos? Serão avaliados eventos de acidentes com responsabilidade do motorista? Serão avaliadas todas as taxas de gravidade de acidentes ou será focado em algum cluster específico?
	Definição da questão de pesquisa	Definir claramente as questões que serão avaliadas na pesquisa; Ex.: Quais as placas ou transportadores geram mais alertas? Quais os transportadores com maior índice de alertas e acidentes? Quais os transportadores têm maior impacto? Qual a sequência de eventos que antecede um acidente? Qual a frequência comum de eventos que antecede um acidente, existe um tempo entre os eventos?
	Selecionar ferramenta de mineração	Definir qual a plataforma de mineração de processos será utilizada, será ferramenta de acesso gratuito? As políticas de privacidade da empresa permitem o acesso à ferramenta? Existe conhecimento técnico disponível para utilização? A ferramenta suporta as análises que constam no objetivo de pesquisa?
	Compor equipe do projeto	Eleger representantes das áreas que irão participar da pesquisa, importante considerar especialistas, técnicos em segurança do trabalho, transportadores, motoristas, equipe de monitoramento e time de tecnologia (TI), gerentes de transporte.
Extração	Identificar origem dos dados	Envolver equipe do projeto para entender quais são os dados disponíveis necessários para responder questões de pesquisa. Avaliar em quais sistemas esses dados estão disponíveis, os dados estão integrados? Avaliar rastreador do veículo, tecnologias de assistência ao motorista, sistema TMS, base histórica de acidentes e incidentes.
	Extrair os dados	Avaliar se equipe de TI consegue extrair log de eventos, caso não seja possível gerar relatórios diretamente dos sistemas, existem APIs disponíveis?

	Transferir conhecimento do processo	Utilizar técnicas como entrevistas, seções de brainstorming de maneira a acelerar entendimento do processo, buscar documentação existente, esse entendimento inicial é fundamental para nortear adequadamente a etapa de pré-processamento dos dados. O que pode impactar na qualidade dos dados? Que mudanças aconteceram recentemente e que podem impactar nos resultados? Toda frota possui equipamentos capazes de gerar alertas?
Pré-Processamento	Definição de campos chaves	Quais os campos disponíveis poderão auxiliar na resposta das questões de pesquisa? Placa, Motorista, viagem? Qual campo será o CASE ID? Condições climáticas na hora do acidente? Velocidade média? Quais os campos podem ser desconsiderados?
	Saneamento dos dados	Remoção de logs incompletos, remoção de cases incompletos, remoção de dados sem placa, avaliar se todos os dados estão no período de análise.
	Formatação e consolidação	Para casos em que os dados não sejam integrados será necessário avaliar e preparar as bases de maneira a garantir base consolidada de eventos onde seja possível identificar claramente as informações de Case ID, <i>Timestamp</i> e Nome do evento. Nesta etapa também é importante preparar os dados para que dados pessoais não sejam expostos, codificar dados de CPF e nome de motorista. Estar atento a Leis de proteção de dados.
Processamento dos dados	Criação da Visualização	Os logs de eventos precisam estar preparados para responder questões de pesquisa. O caso ou estância pode variar de acordo com a pergunta a ser respondida. Ex.: Se a resposta está relacionada ao comportamento dos motoristas ou transportadores estes campos podem ser utilizados com case ID.
	Agregação de eventos	Avaliar eventos que podem agregar nas questões de pesquisa. Dados de histórico de conduta do motorista e transportador.
	Registros enriquecedores	Existem registros que podem melhorar a qualidade do modelo, exemplo uma base de acidentes com dados históricos que possa ser conectada?
	Registros de filtragem	Avaliar uma grande base de dados pode dificultar a visualização das informações, assim pode ser realizar filtros específicos como por exemplo o código de uma transportara, o período específico. O cluster de gravidade de acidente, a responsabilidade do acidente se motorista ou terceiro, o dia de semana específico ou horário do dia.
Mineração e análise	Detecção dos processos	Dado um log de eventos como entrada, iniciar análise com técnicas de descoberta de processo, que retornam um modelo de processo baseado em fatos como saída.
	Verificação de conformidade	Comparar se os eventos gerados estão em desacordo com políticas e regulamentos
	Aprimoramento	Avaliar que respostas podem ser obtidas através dos logs de eventos. Ex.: Quais são principais alertas gerados? A identificação pode auxiliar na evolução dos processos?
	Análise de processo	Aplicar Técnicas de mineração de dados ou análise visual (por exemplo, histogramas de eventos por caso), cujos resultados podem ser usados para aprimorar modelos de processo com aspectos adicionais.

Avaliação	Diagnosticar	Interpretar corretamente os resultados (por exemplo, compreender o modelo de processo descoberto); distinguir interessante ou incomum resultados dos esperados (por exemplo, grande conjunto de execuções anormais) e identificação ou refinamento de questões de pesquisa para iterações posteriores possíveis.
	Verificar e validar	Tanto a verificação quanto a validação podem ajudar a identificar as causas básicas subjacentes e a projetar ideias para possíveis melhorias no processo.

FONTE: O autor (2023).

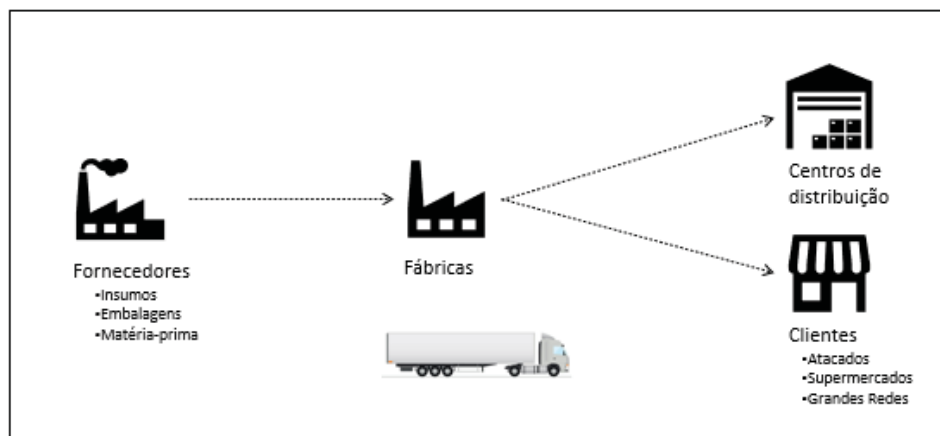
O framework descrito no **Quadro 5** detalha cada fase da mineração de processos e destaca a importância da colaboração entre diferentes áreas, como segurança do trabalho, TI e operações de transporte. Cada etapa foi cuidadosamente planejada para garantir que as questões de pesquisa sejam respondidas de forma clara e que as análises sejam conduzidas com a máxima precisão possível. A fase de extração, por exemplo, é crítica para garantir que os dados necessários estejam acessíveis e em um formato que permita uma análise eficiente. Além disso, as atividades de pré-processamento, como o saneamento e a formatação dos dados, asseguram que as informações relevantes sejam corretamente interpretadas, minimizando a possibilidade de erros.

## 5 APLICAÇÃO

Para validar a efetividade do PM<sup>2</sup>RCS, o framework foi aplicado em um estudo de caso real no transporte de longa distância de uma renomada multinacional do setor alimentício. Escolhida pela facilidade de acesso aos dados e forte compromisso com a segurança e prevenção de acidentes. Este estudo de caso focou especificamente no transporte rodoviário de cargas de longo curso, com o objetivo de garantir uma logística segura e eficiente para o abastecimento de fábricas, centros de distribuição e clientes finais, evidenciando a aplicabilidade prática do PM<sup>2</sup>RCS na melhoria contínua da segurança no transporte de cargas.

A Figura 8 apresenta uma representação gráfica da logística de longo curso adotada pela empresa. Destaca-se que a redução de acidentes figura como um dos principais objetivos da área de transporte, que tem se dedicado à implementação de novas tecnologias embarcadas ao longo dos últimos anos para aprimorar suas operações.

FIGURA 8 – REPRESENTAÇÃO LOGÍSTICA DE LONGO CURSO



FONTE: O autor (2023).

Este estudo de caso busca não apenas examinar como o Framework PM<sup>2</sup>RC se manifesta na prática, mas também entender de que maneira ele contribui para o alcance dos objetivos da empresa em termos de segurança e eficiência no transporte rodoviário para o fluxo apresentado na Figura 8.

O compromisso contínuo com a implementação de tecnologias inovadoras reflete o empenho da organização em promover uma gestão de transporte cada vez mais segura e eficaz.

## 5.1 PLANEJAMENTO

A etapa de planejamento contempla as atividades de Seleção do processo, Definição das questões de pesquisa, seleção da ferramenta de mineração e a composição da equipe do projeto conforme detalhamento a seguir.

### 5.1.1 Selecionar processo

Para definir o escopo do projeto foi realizado alinhamento com Gerente executivo da área de transporte pois manifestou antecipadamente interesse na aplicação de novas técnicas para avaliação de risco de acidentes. A empresa, recentemente assumiu um compromisso público em prol da redução de acidentes, demonstrando dedicação na identificação de tecnologias capazes de contribuir para a prevenção de incidentes. Atualmente, 90% dos veículos já estão equipados com Telemetria, e há um plano em andamento para dotar toda a frota com sensores de fadiga.

### 5.1.2 Compor equipe do Projeto

Na fase de formação da equipe do projeto, optou-se por reunir a equipe de projetos da logística, o coordenador de gestão de risco e os fornecedores de tecnologia. É relevante notar que todos os membros selecionados não possuíam experiência prévia em mineração de processos. Essa decisão estratégica foi guiada pela busca de uma abordagem multidisciplinar e pela diversidade de habilidades, visando enfrentar os desafios específicos do projeto.

Ao compor o time, considerou-se as habilidades individuais de cada membro, visando uma combinação sinérgica de competências. A diversidade de perspectivas e conhecimentos foi encarada como um ativo, contribuindo para uma visão mais abrangente e inovadora na abordagem da mineração de processos.

A formação do time foi conduzida de maneira a endereçar as lacunas de competências identificadas, garantindo que o conjunto de habilidades estivesse alinhado com as demandas específicas do projeto. Essa abordagem estratégica na seleção da equipe busca assegurar a capacidade do time em superar os desafios únicos que a mineração de processos pode apresentar.

Além disso, destaca-se que a falta de experiência prévia em mineração de processos foi encarada como uma oportunidade de aprendizado e crescimento. Estratégias de integração foram implementadas para fomentar a comunicação eficaz, a colaboração e a construção de um ambiente coeso de trabalho, promovendo uma atmosfera propícia à troca de conhecimentos e à superação dos desafios.

### 5.1.3 Definição da questão de pesquisa

Na etapa de definição da questão de pesquisa, foram elaboradas as seguintes indagações para orientar a investigação:

1. Existe um padrão perceptível no comportamento dos motoristas durante a geração de alertas durante as viagens?
2. A quantidade de alertas está correlacionada a um maior risco de acidentes?
3. Qual tipo de alerta indica, de maneira mais substancial, a iminência de um acidente?
4. A utilização da mineração de processos oferece uma maneira eficiente de visualizar dados de rastreamento de veículos, incidentes e outras informações relevantes, facilitando uma análise mais acessível e aprimorada?

Essas questões foram formuladas com o objetivo de aprofundar a compreensão do comportamento dos motoristas em relação aos alertas gerados, explorar possíveis correlações entre a quantidade de alertas e o risco de acidentes, identificar os alertas mais críticos e avaliar a viabilidade da mineração de processos como uma ferramenta efetiva para a análise de dados complexos relacionados ao rastreamento de veículos e incidentes. Essa abordagem estratégica visa não apenas responder às perguntas específicas, mas também contribuir para uma compreensão mais abrangente e informada sobre a segurança veicular e a prevenção de acidentes.

#### 5.1.4 Definição da ferramenta de mineração

Optou-se pela utilização da ferramenta de mineração de processos Disco (FLUXICON, s.d.) devido à familiaridade com sua interface e ao fato de ela permitir o processamento de grandes volumes de dados de maneira eficiente. Embora a ferramenta Celonis (CELONIS, s.d.) ofereça um conjunto mais amplo de funcionalidades, sua versão acadêmica possui uma limitação quanto ao número de linhas que podem ser processadas, restrição que não ocorre com o Disco, tornando esta última a escolha mais adequada para o presente estudo.

## 5.2 EXTRAÇÃO

Nessa etapa, iniciou-se o mapeamento das alternativas de coleta de dados, considerando que um dos maiores desafios é o grande número de fontes de informação. A aplicação da mineração de processos pressupõe a existência de três informações essenciais: o caso, que representa a instância do processo; a atividade, que descreve as tarefas ou operações realizadas; e o *timestamp*, que indica o momento de ocorrência da atividade, conforme descrito por Vander der Heijden (2012).

### 5.2.1 Identificar origem dos dados

Para identificar origem dos dados foram realizadas entrevistas com especialistas da área de projetos e time de operação. Foi constatado que as bases não são integradas e que são extraídas diretamente do site dos fornecedores. Existe uma integração dos dados de telemetria, porém a API envia dados uma vez ao dia.

Foram identificadas sete plataformas para obtenção de dados relacionados aos acidentes conforme detalhado na Tabela 1.

TABELA 1 – FONTE DOS DADOS

Sistema	Informação Extraída
ERP SAP	Viagens realizadas, logs de início e fim carregamento e descarga

Telemetria	Eventos gerados telemetria (4 Fornecedores)
Plataformas fadiga	Eventos gerados sensores de fadiga (3 Fornecedores)
SAI	Dados de acidentes (sistema de investigação de acidentes)
LINCROS	Alertas de viagem – Sistema de monitoramento da frota
TMS	Dados de planejamento da viagem
APISUL	Sistema de gestão de risco
DEMARCO	Prontuário do condutor

---

FONTE: O autor (2023).

### 5.2.2 Extrair os dados

A extração dos dados foi realizada pelo especialista em tecnologia responsável pela área de projetos. Cada base de dados foi individualmente extraída de suas respectivas fontes. A ausência de uma chave comum para associar os eventos em algumas fontes exigiu uma manipulação cuidadosa dos dados para garantir a integridade e a precisão das informações correlacionadas.

A primeira base de dados obtida foi a de registro de acidentes, os registros são controlados em um sistema onde são armazenadas as informações identificadas no processo de investigação de acidentes. A base possui registros desde 2017 de acidentes e incidentes tanto de responsabilidade do motorista (acidentes que durante investigação ficou comprovada a responsabilidade do motorista) quanto de responsabilidade de terceiros (acidentes que durante investigação ficou comprovada responsabilidade do terceiro).

As informações de alertas de telemetria e fadiga foram obtidas de um arquivo utilizado pela equipe de gestão de sinistros. Os arquivos são baixados diariamente e consolidados em arquivos de Excel separados e com formatos diferentes. Também foram baixados arquivos de carregamentos e descargas do ERP. Não foi necessário buscar dados do TMS e do sistema de monitoramento pois as informações puderam ser obtidas direto do ERP. Outra fonte de informação gerada foi o prontuário do condutor, obtida na plataforma de gestão do motorista.

### 5.2.3 Transferir conhecimento do processo

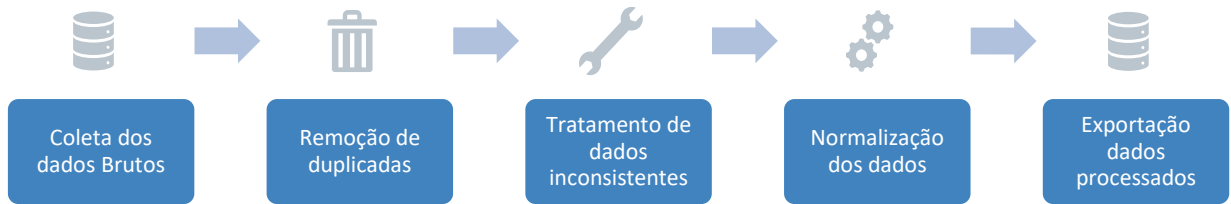
Para obtenção de informações sobre o processo foram realizadas 4 sessões com o time de operações e time de projetos identificando o que já foi aplicado até agora para ter efeito positivo na redução de acidentes, nessa etapa foi verificado que a empresa tem uma preocupação genuína com a garantia da segurança dos motoristas e que já aplicou grande parte das recomendações para redução de acidentes. Além disso também se iniciou a participação nas principais agendas de tratativa do tema na empresa:

- Fórum mensal de segurança no transporte: Evento com participação de 3 Vice-presidentes, Gerentes executivos das áreas de transporte e Recursos humanos e Gerente de Projetos. Neste evento são apresentados indicadores mensais de resultados além das investigações de acidentes ocorridas no mês além de gerar demandas de estudos e novos projetos.
- Agendas de investigação de acidentes: Evento que ocorre em caso de ocorrência de acidentes, neste o transportador apresenta informações sobre o acidente e uma equipe composta por líderes de transporte e especialista em segurança realiza os questionamentos com objetivo de chegar na causa fundamental. Nesta agenda são avaliados indicadores históricos sobre o comportamento do motorista além de visualizar registros como gravações geradas pelo sistema de prevenção de fadiga e alertas gerados pela telemetria.

## 5.3 PRÉ-PROCESSAMENTO

A etapa de pré-processamento foi uma das mais desafiadoras do projeto, sendo crucial para garantir a integridade e a consistência das informações antes da aplicação da metodologia de mineração de processos. O fluxo de etapas realizado pode ser visualizado na Figura 9, que apresenta desde a coleta dos dados brutos até a exportação dos dados processados, passando pela remoção de duplicatas, tratamento de inconsistências e normalização.

FIGURA 9– ETAPAS DO PREPROCESSAMENTO



Fonte: O Autor (2023)

A base de acidentes utilizada contém um histórico de mais de 10 anos de ocorrências, com mais de 40 campos detalhando cada evento. No entanto, uma das principais limitações dessa base era a ausência de um campo que conectasse essas informações a outras bases, como as de sinistros. Para resolver esse problema, foi necessário realizar a integração com a base de sinistros, utilizando o número de sinistro como chave primária para relacionar as duas bases. Incidentes que não possuíam essa correlação foram desconsiderados, pois não havia como conectá-los adequadamente, o que poderia comprometer a análise.

As bases de dados de carregamento e descarga foram geradas por relatórios de tempos de processo, que não estavam no formato adequado para a aplicação das técnicas de Mineração de Processos. Essas bases não possuíam uma “noção de case” — ou seja, não havia uma estrutura clara que identificasse o fluxo completo de cada processo, essencial para a análise de processos (van der AALST, 2011).

As bases de carregamento e descargas foram geradas por um relatório de tempos de processo e dessa foram não estavam no formato adequado para a mineração (sem noção de case). Para transformar a base em uma base adequada para mineração (com noção de case) e realizar o saneamento da base foi utilizado código em Python. Nesta etapa foram tratados dados faltantes realizado a transformação de linhas em colunas conforme código apresentado na Figura 10.

FIGURA 10– FORMATO DADOS SEM INCLUSÃO CASE

Viagemx	Atividade1	Timestamp1	Atividade2	Timestamp2	Atividade3	Timestamp3	Atividade4	Timestamp4
---------	------------	------------	------------	------------	------------	------------	------------	------------

Fonte: O autor (2023)

Para transformar esses dados em uma base adequada para a mineração, foi utilizado um algoritmo em Python, que permitiu a transformação de linhas em colunas,

criando a noção de case necessária. Além disso, foi realizado o tratamento de dados faltantes e a normalização das variáveis conforme apresentado no Código 1.

### Código 1 – Algoritmo em Python para Pré-processamento de Dados

```
import pandas as pd.
import numpy as np
from sklearn.preprocessing import MinMaxScaler

# Carregar os dados brutos
df = pd.read_csv('dados_carregamento.csv')

# Remover duplicatas
df = df.drop_duplicates()

# Preencher valores ausentes com a média das colunas relevantes
df.fillna(df.mean(), inplace=True)

# Normalização de colunas específicas: velocidade e tempo de atividade
scaler = MinMaxScaler()
df[['velocidade', 'tempo_atividade']] =
scaler.fit_transform(df[['velocidade', 'tempo_atividade']])

# Transformar dados de linhas para colunas (para criar noção de case)
df_transformado = df.pivot(index='viagem', columns='atividade',
values='timestamp')

# Exportar os dados ajustados
df_transformado.to_csv('dados_preprocessados.csv')
```

Fonte: O Autor 2023

A seguir são detalhadas as etapas apresentadas no Código 1.

1. Carregamento dos dados: Os dados foram carregados em um DataFrame do Pandas (MCKINNEY, 2010).
2. Remoção de duplicatas: Foram eliminados registros duplicados para evitar distorções nas análises futuras.
3. Tratamento de dados faltantes: Utilizou-se a média das colunas para preencher os valores ausentes, uma técnica recomendada para evitar a perda de grandes quantidades de dados (JAMES et al., 2013).
4. Normalização: Aplicou-se o MinMaxScaler da biblioteca Scikit-learn para ajustar os valores de variáveis como velocidade e tempo de atividade a uma escala

comum, entre 0 e 1, para garantir que todas as variáveis fossem comparáveis entre si (PEDREGOSA et al., 2011).

5. Transformação dos dados: A função pivot() foi utilizada para reorganizar os dados de linhas para colunas, criando a noção de case necessária para a análise de Mineração de Processos.

Após passar pela etapa de saneamento foi gerada a base de dados com noção de case, também foram mantidas apenas colunas que seriam utilizadas na mineração. Esta etapa foi repetida a cada interação onde se percebeu necessidade de novos campos para análise. Na figura 11 é possível evidenciar como ficou a base de dados ajustada.

FIGURA 11 – BASE DE DADOS AJUSTADA COM INCLUSÃO DE CASE

	Load	variable	value
0	2519023	Inicio Carregamento	2022-02-01 00:24:00
1	2515412	Inicio Carregamento	2022-02-01 00:32:26
2	2518067	Inicio Carregamento	2022-02-01 00:51:23
3	2517126	Inicio Carregamento	2022-02-01 01:02:11
4	2517476	Inicio Carregamento	2022-02-01 02:15:55

Fonte: O Autor (2023)

Os dados de telemetria e fadiga também precisaram ser consolidados. Primeiramente foram consolidados os arquivos de telemetria unindo as bases mensais e da mesma forma para os dados de fadiga. Na Figura 12 é possível ver um exemplo das bases de eventos gerados pela tecnologia de telemetria separados por mês em arquivo de Excel.

FIGURA 12 – LISTA DE ARQUIVOS DADOS DE EVENTOS TELEMETRIA E FADIGA

Nome	↓
<input checked="" type="checkbox"/>	Consolidado 31.05.xlsx
<input checked="" type="checkbox"/>	Consolidado 31.03.xlsx
<input checked="" type="checkbox"/>	Consolidado 31.01.xlsx
<input checked="" type="checkbox"/>	Consolidado 30.04.xlsx
<input checked="" type="checkbox"/>	Consolidado 28.06.xlsx
<input checked="" type="checkbox"/>	Consolidado 28.02.xlsx

FONTE: O Autor (2023)

Para garantir que os dados de telemetria fossem processados de forma eficiente e estivessem prontos para a análise de Mineração de Processos, foi utilizado um algoritmo em Python para consolidar múltiplos arquivos Excel que continham dados coletados mensalmente.

O código Python permitiu a automatização da tarefa de leitura e concatenação dos arquivos de forma a garantir que todas as informações de diferentes meses fossem unificadas corretamente em um único conjunto de dados. O Código 2 ilustra como os arquivos foram consolidados:

### Código 2 – Algoritmo em Python para Pré-processamento de Dados

```
import os
import pandas as pd

# Definir o diretório onde os arquivos estão localizados
diretorio = 'caminho/para/seu/diretorio'

# Inicializar uma lista para armazenar os dataframes de cada arquivo
consolidado_list = []

# Iterar sobre os arquivos no diretório
for filename in os.listdir(diretorio):
    if filename.startswith('Consolidado') and filename.endswith('.xlsx'):
        file_path = os.path.join(diretorio, filename)
        df = pd.read_excel(file_path)
        consolidado_list.append(df)

# Concatenar todos os dataframes da lista
df_consolidado = pd.concat(consolidado_list, ignore_index=True)

# Exportar o dataframe consolidado
df_consolidado.to_excel ('caminho/para/salvar/Consolidado_Final.xlsx',
index=False).
```

Fonte: O Autor (2023)

Com o código apresentado, foi possível consolidar todos os arquivos mensalmente separados em um único arquivo Excel, facilitando o pré-processamento e a análise subsequente. Esse processo automatizado reduziu significativamente o tempo de manipulação manual de dados.

## 5.4 PROCESSAMENTO DOS DADOS

A etapa de processamento dos dados, teve como foco a transformação e estruturação dos logs de eventos, organizando-os de maneira a facilitar a análise. Após o pré-processamento, os dados de acidentes passaram por um refinamento adicional, com a criação de visões, agregação de eventos e enriquecimento dos logs.

Na primeira fase do processamento, foi realizada a criação de visões, onde os dados de acidentes foram organizados em torno de casos específicos (instâncias de acidentes) e suas respectivas classes de eventos (como eventos de telemetria, eventos de fadiga e atividades associadas ao motorista). Essa estruturação foi fundamental para permitir a análise dos acidentes ao longo do processo e identificar padrões. Para facilitar a visualização a base de dados foi separada em duas, uma com registros de viagens com acidentes e outra com registros de viagens sem acidentes.

A próxima etapa consistiu no estabelecimento dos registros de filtragem, foi definido em conjunto com operação e especialista da área de projetos avaliar apenas eventos de viagens que depois da investigação ficou constatada responsabilidade do motorista.

O passo final no processamento foi o enriquecimento dos logs. Aqui, foram incorporadas informações adicionais, como idade do motorista, dados sobre as condições climáticas no momento do acidente, e outros fatores que poderiam influenciar o processo de análise.

Nessa etapa de processamento de dados, o feedback da especialista do negócio com relação às conclusões intermediárias sobre os dados foi frequente (na maior parte das vezes, utilizando-se de imagens e desenhos). Com isso foi possível inicializar e atualizar o processamento de dados após cada iteração.

## 5.5 MINERAÇÃO E ANÁLISE

A base de dados de acidentes utilizada nesta pesquisa fornece um amplo panorama dos acidentes envolvendo veículos de transporte em um contexto específico. Abrangendo um período que vai de Jan/2020 a Dez de 2023, esta base é composta por registros detalhados de cada acidente, onde cada linha representa um incidente distinto são 32 colunas que podem ser usadas no processo de mineração. No Quadro 9 pode ser vista a descrição de cada campo.

Para a análise foram considerados dados gerados a partir de 2020, essa definição foi realizada em conjunto com equipe do projeto por entender que a partir

deste período os dados estão mais robustos e confiáveis já que muitas variáveis foram sendo incluídas nas investigações ao longo dos anos.

QUADRO 3 – DESCRIÇÃO CAMPOS TABELA DE ACIDENTES

<b>Nome Campo</b>	<b>Descrição</b>	<b>Tipo Variável</b>
Nº Simplificado	Número do registro do acidente	Identificador
Data da ocorrência do Sinistro	Data de ocorrência do acidente	Data
Hora_Acidente	Hora que ocorreu o acidente	Data
Tipo de Acidente	Classificação do acidente	Categórica
Nº de vítimas fatais	Quantas pessoas faleceram	Numérica
Ação Direta	Fala sobre responsabilidade do acidente	Categórica
Segmento	Tipo de carga	Categórica
Veículo carregado ou vazio	indica se o veículo estava ou não carregado	Categórica
Classificação do Acidente	Classificação do acidente	Categórica
Placa cavalo / Truck	placa do veículo envolvido no acidente	Identificador
Transportadora frota ou eventual	dado indicativo sobre relação contratual com frota	Categórica
Idade Motorista	qual a idade do condutor	Numérica
Estado Civil	estado civil do condutor	Categórica
Direção Defensiva Válida	diz respeito a treinamentos realizados pelo motorista, se estava vencido ou dentro da validade	Categórica
Tempo de Habilitação	Tempo que o motorista possui habilitação	Numérica
CNH Válida	A carteira de habilitação estava valida	Categórica
Possui Multas CNH	o condutor possui multas	Categórica
Conhecia rota	o condutor já conhecia a rota	Categórica
Problemas familiares	No momento do acidente o motorista tinha algum problema familiar	Categórica
Faz uso de medicamentos	o condutor fazia uso de medicamentos	Categórica
Alcoolizado	Motorista estava alcoolizado	Categórica
Tempo de empresa(meses)	Tempo que motorista trabalha no transportador atual	Numérica
Notificações SSMA	Diz respeito a problemas comportamentais anteriores como excesso de velocidade, falta de uso de cinto de segurança	Categórica
Tempo que dirige o modelo de cavalo atual (meses)	Tempo que dirige o tipo de veículo atual, diz respeito a familiaridade com a direção do tipo de veículo	Numérica
Velocidade momento do acidente	Velocidade que estava no momento do acidente	Numérica
Jornada trabalho foi respeitada (intra jornada e direção diária)	foi respeitada jornada de trabalho	Categórica
Traçado	como era o traçado da rodovia	Categórica
Estado conservação	diz respeito a conservação da via	Categórica

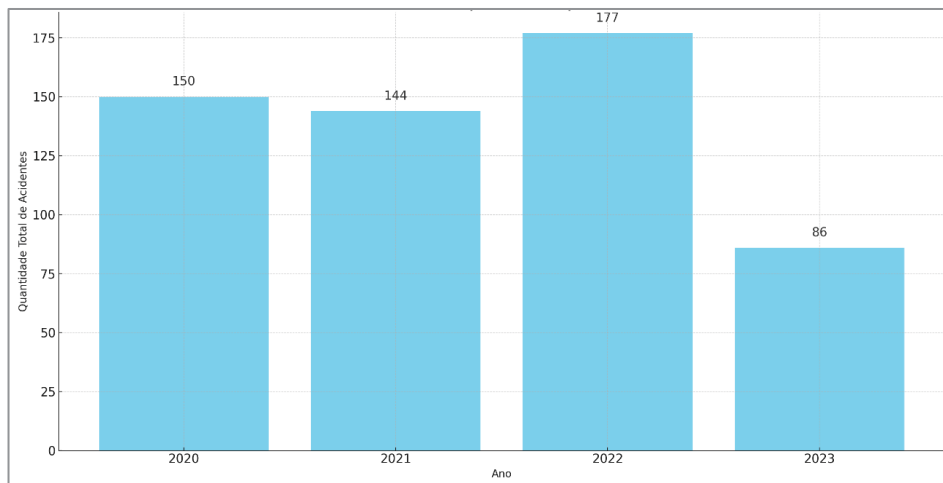
Pista	diz respeito a complexidade da direção, pista era reta ou curva	Categórica
Velocidade da Via	qual era a velocidade permitida na via	Numérica
Condições climáticas	como estava o clima	Categórica
Idade Cavalo	qual a idade do veículo	Numérica

Fonte: O autor (2023)

### 5.5.1 Mineração dados Acidentes

Para iniciar o entendimento dos dados foi realizada uma análise exploratória dos dados iniciando pela evolução da quantidade de acidentes ao longo do ano conforme apresentado no Gráfico 2.

GRÁFICO 2 – EVOLUÇÃO NA QUANTIDADE DE ACIDENTES (2020 a 2023)



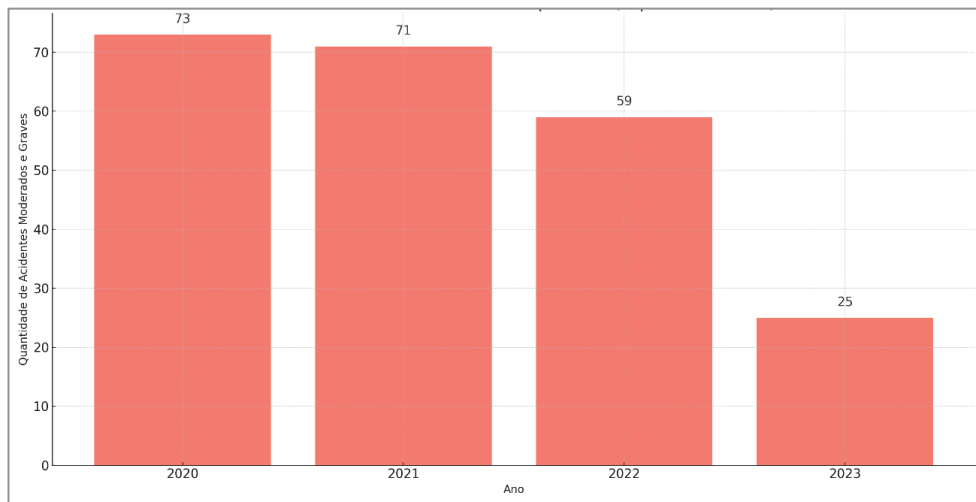
Fonte: O Autor (2024)

O Gráfico 2, indica uma variação na quantidade total de acidentes ao longo dos anos considerados. Este padrão pode refletir a eficácia das medidas de segurança implementadas, mudanças no volume de transporte, ou outros fatores externos que influenciam a ocorrência de acidentes. A visão anual proporcionada pelo gráfico é ajuda a entender a eficácia das políticas de segurança ao longo do tempo e identificar anos que necessitam de análises mais aprofundadas para entender as razões por trás do aumento ou diminuição dos acidentes.

Na sequência foram avaliadas as informações referente à classificação de acidentes moderados a grave.

O Gráfico 3 fornece um olhar focado em casos mais sérios, que são de particular interesse devido ao seu potencial impacto em termos de perdas humanas e materiais. A análise deste gráfico apoia na avaliação da severidade dos acidentes ocorridos e para identificar tendências ou padrões específicos em acidentes graves, possibilitando a implementação de medidas preventivas mais direcionadas.

GRÁFICO 3 – EVOLUÇÃO NA QUANTIDADE DE ACIDENTES MODERADOS e Graves (2020 a 2023)

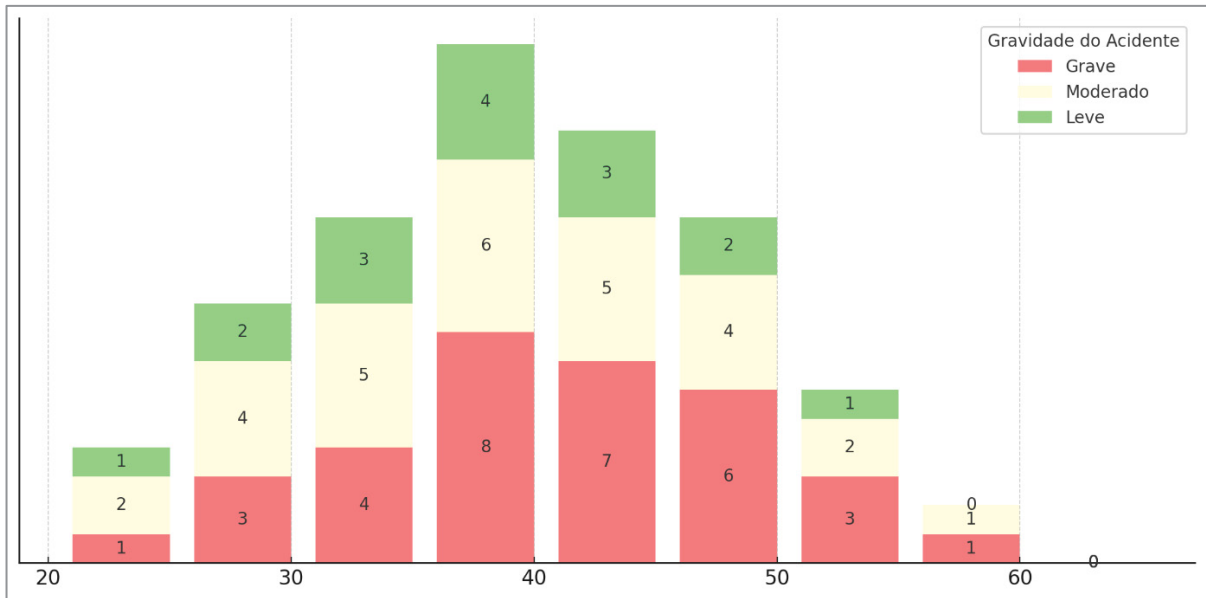


Fonte: O Autor (2024)

O Gráfico 3 apresenta a evolução dos acidentes moderados e graves de 2020 a 2023, com uma tendência clara de redução. Em 2020, houve 73 acidentes, enquanto em 2023 esse número caiu para apenas 25. Esse declínio representa uma queda de aproximadamente 66%, o que reflete os esforços significativos em medidas de segurança adotadas pela empresa.

O próximo passo foi realizar uma análise demográfica dos acidentes com entendimento das principais características. O Gráfico 4 apresenta a distribuição dos acidentes x idade dos motoristas. O Gráfico 4 apresenta a distribuição de acidentes por diferentes faixas etárias de motoristas, destacando a frequência e gravidade dos acidentes (Leve - Verde, Moderado - Amarelo, Grave - Vermelho). Essa visualização permite identificar se determinadas faixas etárias estão mais suscetíveis a se envolver em acidentes e se a gravidade desses eventos varia com a idade do condutor.

GRÁFICO 4 – EVOLUÇÃO NA QUANTIDADE DE ACIDENTES X GRAVIDADE (2020 a 2023)

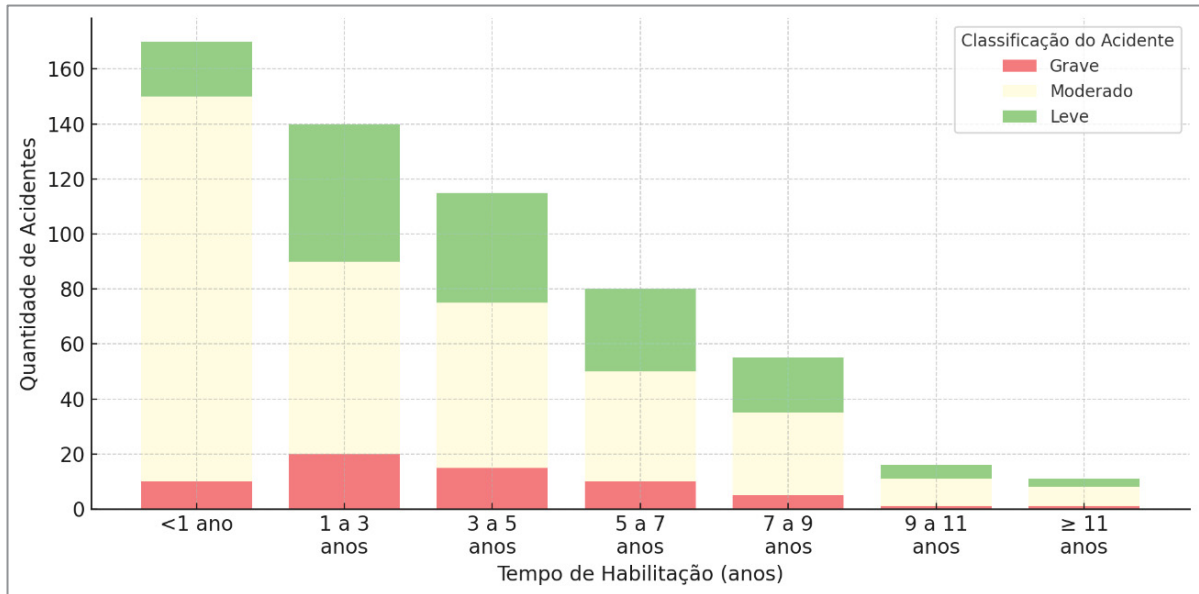


Fonte: O Autor (2024)

Os dados apontam uma correlação entre a idade do motorista e a gravidade dos acidentes, sugerindo que faixas etárias mais jovens tendem a registrar uma maior proporção de acidentes moderados e graves. Em contrapartida, motoristas mais experientes apresentam uma maior frequência de acidentes leves, possivelmente devido à maior habilidade em evitar situações críticas, mas ainda expostos a incidentes de menor gravidade. Essa informação pode orientar as políticas de segurança no transporte rodoviário, especialmente em relação à customização de medidas preventivas por faixa etária.

Outra avaliação realizada foi a de acidentes x tempo de habilitação conforme apresentado no Gráfico 5 que apresenta a relação entre o tempo de habilitação dos motoristas e a quantidade de acidentes, segmentados por gravidade (leve, moderado e grave). A análise dos dados revela que motoristas com menos de 1 ano de habilitação são responsáveis pela maior quantidade de acidentes, com mais de 160 incidentes registrados, sendo a maior parte classificados como leves ou moderados, mas com uma proporção significativa de acidentes graves. Conforme o tempo de habilitação aumenta, há uma tendência de redução no número total de acidentes, com motoristas mais experientes (aqueles com 9 anos ou mais de habilitação) apresentando um número significativamente menor de ocorrências, tanto leves quanto graves. Isso indica que a inexperiência é um fator de risco relevante no aumento de acidentes.

GRÁFICO 5 – DISTRIBUIÇÃO DO TEMPO DE HABILITAÇÃO DOS MOTORISTAS COM ACIDENTES (2020 a 2023)



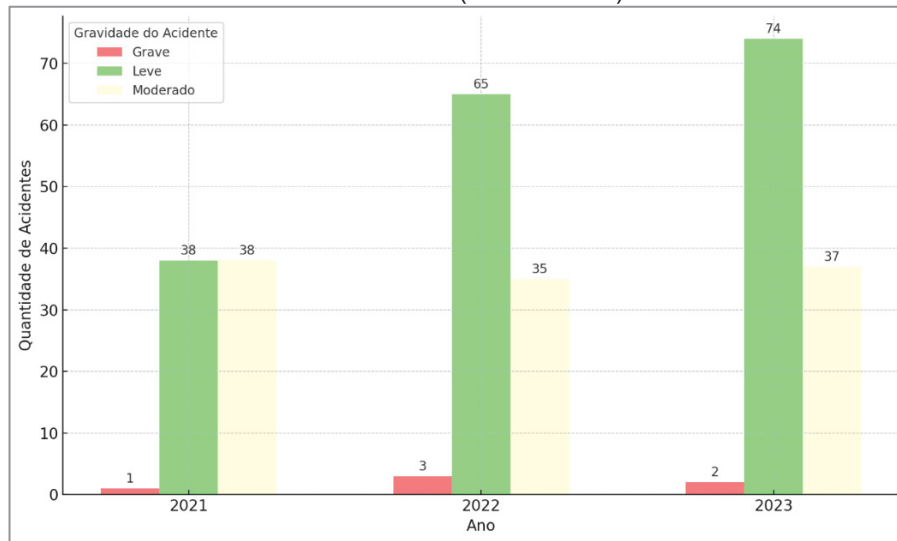
Fonte: O Autor (2024)

Outro ponto importante é que motoristas com 1 a 3 anos de habilitação também apresentam um número expressivo de acidentes, embora com menos casos graves quando comparados aos motoristas com menos de 1 ano de habilitação. Essa diminuição na severidade pode indicar que, com a experiência acumulada, os motoristas desenvolvem habilidades para evitar acidentes mais graves, embora ainda estejam sujeitos a acidentes leves e moderados.

A redução constante de acidentes com o aumento da experiência pode estar associada ao maior conhecimento das estradas, ao desenvolvimento de comportamentos defensivos ao volante e à familiaridade com o veículo e condições de tráfego.

O Gráfico 6 apresenta uma tendência de crescimento nos acidentes leves por responsabilidade do motorista, saindo de 38 em 2021 para 74 em 2023. Já os acidentes moderados e graves mantiveram as quantidades, com pequena variação a menor. A exclusão dos acidentes ocasionados por terceiros concentra a análise nos eventos onde o motorista é o agente direto, oferecendo uma perspectiva mais específica sobre o comportamento do condutor e as circunstâncias do acidente. Este foco é particularmente útil para investigar a eficácia de treinamentos de motoristas, campanhas de conscientização, e outras medidas direcionadas à redução da incidência e gravidade dos acidentes de trânsito.

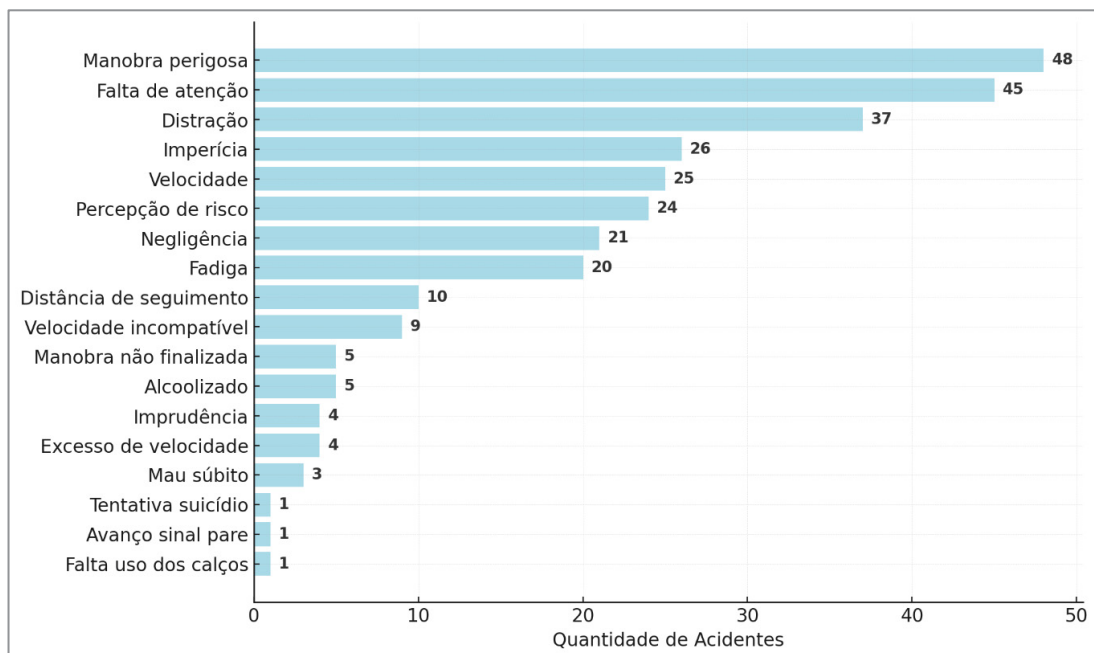
GRÁFICO 6 – QUANTIDADE DE ACIDENTES POR ANO E GRAVIDADE RESPONSABILIDADE MOTORISTA (2020 a 2023)



Fonte: Autor (2024)

Por fim é realizada a avaliação das causas de acidentes apresentado no Gráfico 7. As causas constam na base de dados e são alimentadas depois de haver uma investigação criteriosa. São avaliados relatos da Polícia Rodoviária Federal, Vídeos disponíveis (caso veículo seja equipado com câmera) e dados de telemetria.

GRÁFICO 7 – FREQUÊNCIA DAS CAUSAS DE ACIDENTES RESPONSABILIDADE MOTORISTA (2020 a 2023)



Fonte: O Autor (2024)

O Gráfico 7 oferece uma visão quantitativa e agrupada dos principais fatores que contribuem para a ocorrência de acidentes. A análise revela que 62% dos acidentes estão associados a comportamentos inadequados dos motoristas, como manobras perigosas, falta de atenção, distração, imperícia e excesso de velocidade. Esses comportamentos indicam um padrão de imprudência que contribui significativamente para os acidentes.

Entre as causas comportamentais, a distração e a falta de atenção se destacam, evidenciando que a perda de foco durante a condução é um fator crítico. Além disso, problemas relacionados à fadiga e negligência também aparecem com frequência relevante, reforçando a importância de abordar essas questões para a prevenção de acidentes. Os dados sugerem a necessidade de intensificar programas de conscientização voltados para a atenção ao volante, além de promover iniciativas que reduzam as distrações e incentivem práticas de direção defensiva. A categoria 'excesso de velocidade/incompatibilidade' surge como uma causa frequente, destacando o perigo do não cumprimento dos limites de velocidade. Isso sugere a eficácia de reforçar o controle de velocidade e de promover campanhas educativas sobre os riscos associados à condução em alta velocidade e de potencializar o uso das tecnologias de telemetria e fadiga como aliados para prevenção de acidentes.

A normalização dos dados no gráfico é uma forma de proporcionar uma compreensão clara e unificada das causas dos acidentes. Isso enfatiza a importância de ter dados bem organizados e categorizados corretamente para análises precisas e conclusões confiáveis. Outra abordagem foi clusterização para entendimento dos principais focos de atuação conforme demonstrado na Tabela 2.

TABELA 2 – CLUSTERIZAÇÃO CAUSAS DE ACIDENTES

Cluster	Porcentagem (%)
Outros	36.21
Imprudência/Imperícia	26.90
Distração/Falta de Atenção	15.52
Excesso de Velocidade/Incompatibilidade	13.10
Problemas de Saúde/Condição do Motorista	8.28
<b>Total</b>	<b>100</b>

Fonte: O Autor (2024)

A Tabela 1 apresenta uma visão clara e quantitativa sobre os fatores que contribuíram para esses incidentes. Ao organizar as causas em categorias

específicas, a tabela ajuda a identificar padrões e tendências importantes de forma mais acessível e compreensível.

O cluster "Outros" compõe a maior proporção de acidentes (36.21%), indicando uma ampla gama de causas menos comuns ou específicas que contribuem para acidentes. Isso reflete a complexidade dos fatores de risco no trânsito e a necessidade de estratégias de prevenção abrangentes.

O segundo maior cluster, "Imprudência/Imperícia" (26.90%), destaca a relevância do comportamento de risco dos motoristas nos acidentes. Essa categoria inclui ações como imprudência e imperícia, sugerindo a importância de focar em educação e treinamento de motoristas para melhorar habilidades de direção e consciência de segurança.

Representando 15.52% dos acidentes, o cluster "Distração/Falta de Atenção" destaca a importância de medidas para manter a atenção dos motoristas na estrada, ressaltando os perigos da distração ao volante.

O cluster "Excesso de Velocidade/Incompatibilidade" (13.10%) confirma o risco associado ao excesso de velocidade, apontando para a necessidade de reforçar as medidas de controle de velocidade e conscientizar os motoristas sobre os perigos de não respeitar os limites de velocidade.

Problemas de Saúde e Condição do Motorista: O cluster "Problemas de Saúde/Condição do Motorista" (8.28%) abrange causas relacionadas à condição física e mental dos motoristas, como fadiga e mal súbito, indicando a necessidade de políticas e práticas que considerem a saúde e o bem-estar dos motoristas.

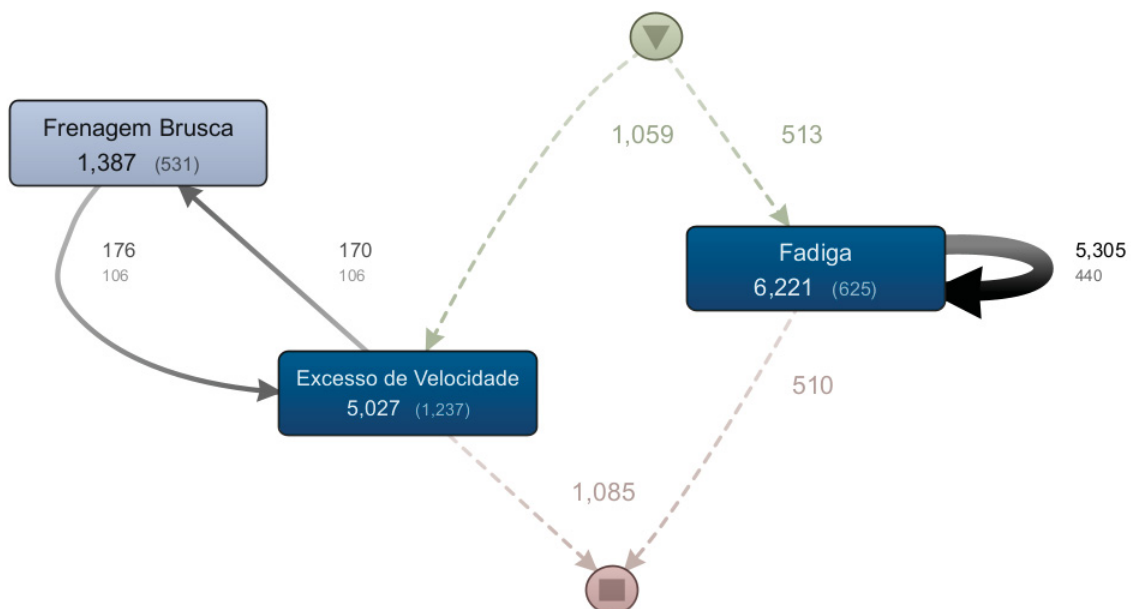
### 5.5.2 Aplicação da Mineração de Processos

Para aplicar a Mineração de Processos (MP), foi adotada a estratégia de segmentar a análise em dois grupos: motoristas que se envolveram em acidentes e motoristas que não se envolveram em acidentes. O objetivo dessa abordagem é identificar possíveis diferenças de comportamento entre os grupos. Como o número de eventos envolvendo motoristas com acidentes é significativamente menor em comparação à base de dados geral, esses eventos tenderiam a ser ofuscados se fossem analisados em conjunto com todos os dados.

Um desafio adicional foi o grande volume de dados, o que inviabilizou a análise completa de um ano inteiro de eventos. No entanto, reduzir o período de análise, especialmente no caso dos acidentes, resultaria em uma base de dados de acidentes muito pequena, o que poderia comprometer a robustez da análise. Para superar essa limitação, optou-se por trabalhar com um ano de eventos relacionados a acidentes e três meses de eventos para motoristas sem acidentes.

Na etapa de mineração de processos, utilizou-se a ferramenta Disco ("Site Fluxicon - Disco", [s.d.]), que gerou um mapa de processos ilustrando o fluxo dos eventos ocorridos durante as viagens. A linha pontilhada no mapa representa o fluxo de eventos ao longo de várias viagens, permitindo a visualização de padrões ou sequências comuns. A análise desses fluxos pode ser crucial para identificar áreas críticas que necessitam de intervenção, como a implementação de programas de segurança focados na redução de comportamentos de risco, como o excesso de velocidade e a fadiga ao volante. A Figura 13 ilustra de maneira clara como os dados foram dispostos e facilita a identificação de padrões relevantes no processo. Essa visualização torna possível uma análise mais detalhada, proporcionando insights para intervenções operacionais e de segurança.

FIGURA 13 – APRESENTAÇÃO FLUXO DOS EVENTOS NO DISCO



FONTE: O Autor (2023)

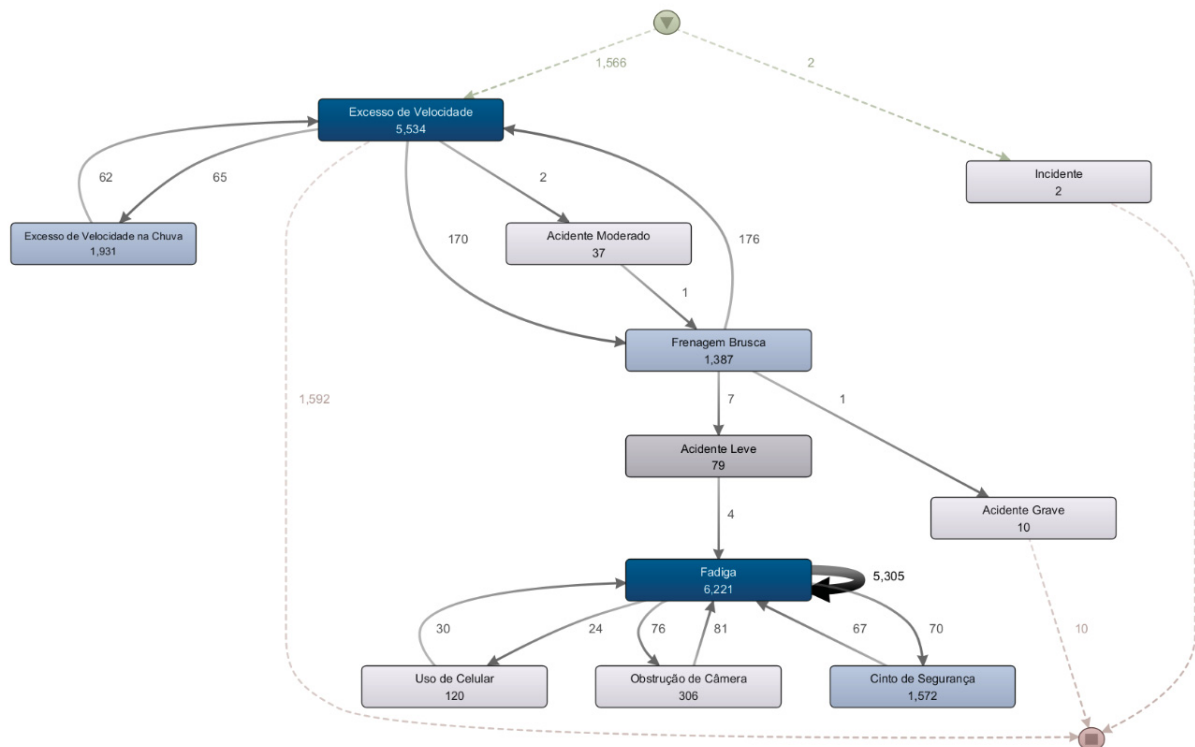
No fluxo de processos, o número próximo à linha pontilhada indica a quantidade de casos analisados, enquanto os números dentro das caixas

representam a quantidade de eventos gerados. As cores mais intensas no mapa indicam os eventos de maior ocorrência.

### 5.5.2.1 Mineração base de eventos com acidentes

Nesta etapa da pesquisa, foi realizada a mineração da base de dados contendo registros de eventos relacionados a motoristas que se envolveram em acidentes. O objetivo desta análise é identificar padrões de comportamento e eventos recorrentes que possam estar associados a esses acidentes, utilizando técnicas de mineração de processos para explorar e correlacionar os dados. A base de dados inclui informações detalhadas sobre as viagens, os motoristas, as placas dos veículos e os diferentes tipos de eventos que ocorreram durante o período de análise. A seguir são apresentados os principais aspectos da base de dados e as inferências obtidas a partir da ferramenta Disco, com foco nas correlações entre os eventos e os acidentes registrados conforme demonstrado na Figura 14.

FIGURA 14 – ANÁLISE DE VARIANTES



Fonte: O Autor (2023)

Conforme apresentado na Figura 14 O Excesso de Velocidade é um evento central, com 5.534 casos registrados, e está diretamente relacionado a outros eventos

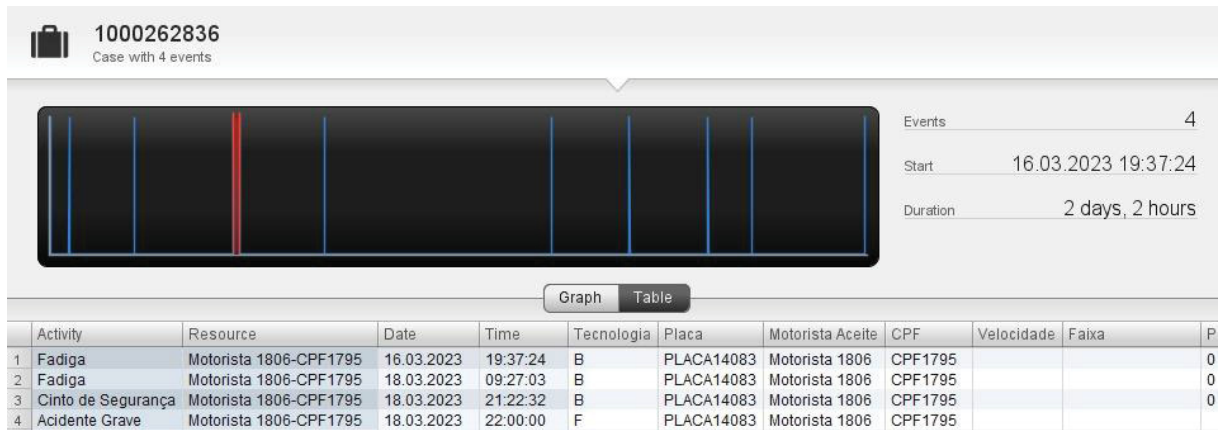
e à ocorrência de acidentes. Em condições climáticas adversas, como a chuva, o excesso de velocidade foi registrado em 1.931 casos, sugerindo um risco elevado nessas situações. Além disso, houve 1.387 registros de Frenagem Brusca, frequentemente resultantes do excesso de velocidade.

Com relação aos acidentes, o fluxo mostra conexões diretas entre o excesso de velocidade e diferentes níveis de gravidade de acidentes. Foram identificados 37 casos de Acidente Moderado, 79 de Acidente Leve e 10 de Acidente Grave, sugerindo uma correlação entre a gravidade dos acidentes e o descontrole de velocidade. A Fadiga é outro fator de destaque, com 6.221 casos registrados, sendo um dos principais fatores relacionados a múltiplos desfechos de acidentes, abrangendo todos os níveis de gravidade. Isso reforça a importância de considerar a fadiga como um elemento crucial na prevenção de acidentes.

O Uso de Celular foi registrado em 120 casos e aparece relacionado à fadiga, sugerindo que o uso do dispositivo contribui para a distração do motorista e aumenta o risco de cansaço. Outro evento significativo é a Obstrução de Câmera, com 306 ocorrências, um fator que pode estar associado à falta de visibilidade ou falhas de equipamento, elevando o risco de acidentes. Por fim, o Cinto de Segurança foi registrado em 1.572 casos e está frequentemente relacionado à fadiga, o que sugere que motoristas fatigados têm menos propensão a utilizar o cinto, agravando as consequências dos acidentes.

O fluxo também revela os caminhos que levam de eventos, como Frenagem Brusca e Fadiga, para diferentes tipos de acidentes. Esses padrões podem ser utilizados para identificar comportamentos de risco e desenvolver intervenções preventivas. Por exemplo, 170 casos de excesso de velocidade resultaram em frenagens bruscas, e dessas, 37 culminaram em acidentes moderados. Esses números ao longo das conexões indicam quantos casos ou viagens seguiram uma determinada sequência de eventos, proporcionando uma visão detalhada dos comportamentos de risco que contribuem para os acidentes. Na sequência seguiu-se a análise com análise de variantes conforme Figura 15.

FIGURA 15 – ANÁLISE DE VARIANTES



FONTE: O autor (2023)

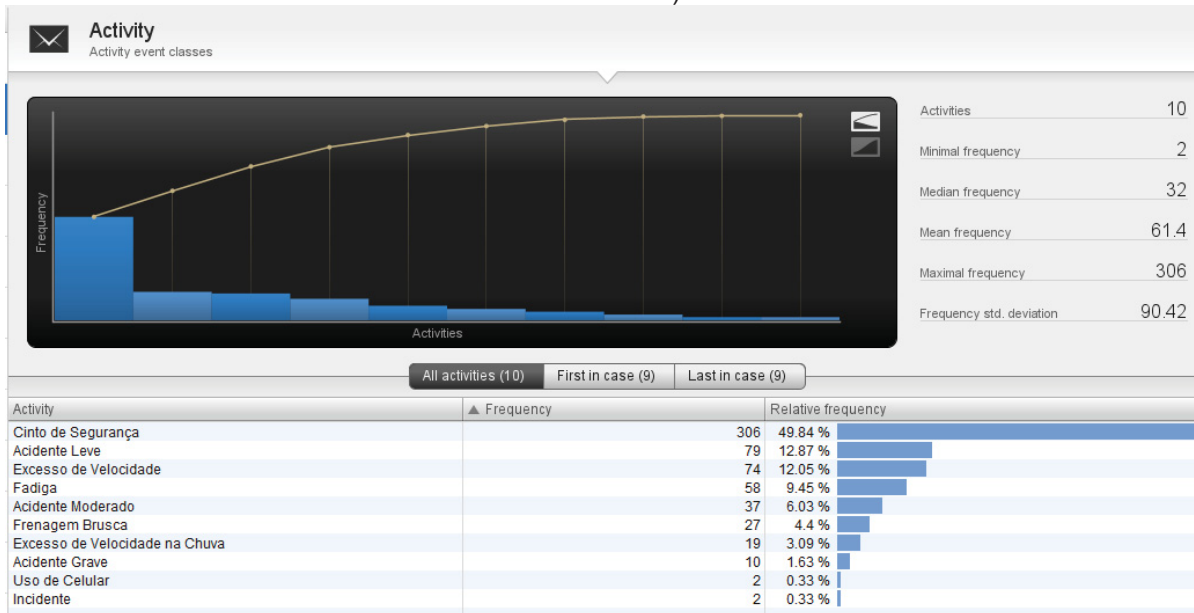
A Figura 15 apresenta um estudo de caso do exemplo número 1000262836 (correspondente ao código da viagem), com o registro de 4 eventos ao longo de 2 dias e 2 horas, permitindo a análise de padrões de comportamento do motorista. O caso começou em 16 de março de 2023, às 19:37:24, com o registro de dois eventos de fadiga nos dias 16 de março e 18 de março de 2023, além de um evento relacionado ao não uso do cinto de segurança no dia 18 de março, às 21:22:32.

O último evento, um acidente grave, ocorreu no dia 18 de março de 2023, às 22:00:00, sendo o fechamento desta série de eventos. Todos os incidentes foram associados ao mesmo motorista, identificado como Motorista 1806-CPF1795. A interpretação dos dados revela ocorrência de fadiga detectado em dois momentos diferentes, seguido de um comportamento de risco adicional (não utilização do cinto de segurança) logo antes do acidente. A combinação desses fatores pode ter contribuído diretamente para a ocorrência do acidente grave. Conforme especialistas em segurança, incidentes de fadiga são considerados um dos principais indicadores de risco para acidentes graves, especialmente quando associados a outras infrações de segurança, como o não uso do cinto ((WIERWILLE, 1999; MAY; BALDWIN, 2009; DINGUS et al., 2016).

A Figura 16, gerada também na ferramenta de mineração apresenta um resumo das principais causas de acidentes ocorridas no período. Analisando as informações pode-se verificar que há 10 atividades diferentes registradas. *Minimal* frequency (Frequência mínima), a atividade menos frequente ocorreu duas vezes. A mediana das frequências é 32, o que significa que metade das atividades ocorreu 32 vezes ou menos, e a outra metade ocorreu mais de 32 vezes. Em média, uma atividade ocorreu aproximadamente 61.4 vezes. A atividade mais frequente ocorreu 306 vezes. O desvio

padrão da frequência é 90.42, indicando a variação da frequência das atividades em relação à média.

FIGURA 16 – ANÁLISE ESTATÍSTICA FREQUÊNCIA EVENTOS VIAGENS COM ACIDENTES (Jan a dez/2023)

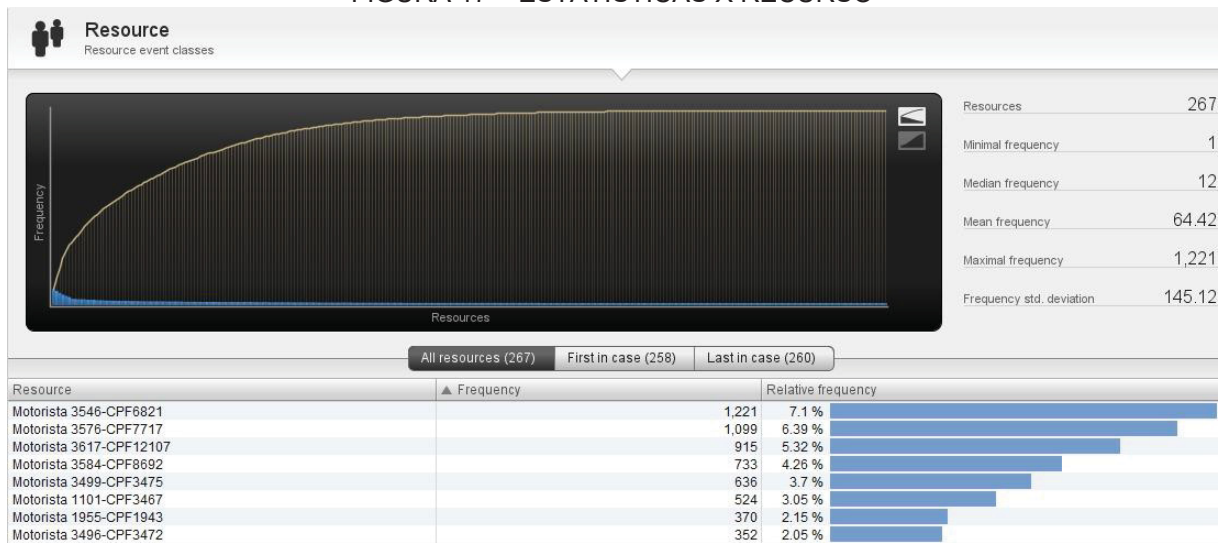


Fonte: O autor (2023)

O alerta de Cinto de Segurança (falta de uso de cinto de segurança) é a atividade mais comum, ocorrendo 306 vezes, o que representa quase metade (49.84%) de todas as atividades registradas. Atividades como Uso de Celular e Incidente são as menos frequentes, cada uma ocorrendo apenas 2 vezes, o que corresponde a 0.33% das atividades. Excesso de Velocidade e Fadiga também são relativamente comuns, com 74 e 58 ocorrências, respectivamente.

Outra análise foi realizada com base na frequência de eventos por recurso conforme apresentado na Figura 17, que exibe a frequência de eventos relacionados a diferentes motoristas, representados de forma codificada. Cada barra no gráfico corresponde a um motorista, e a altura da barra indica a quantidade de eventos associados a esse motorista específico. Ao todo, há 267 motoristas diferentes registrados no conjunto de dados. A figura também apresenta várias estatísticas sobre a distribuição dos eventos, como a frequência mínima de 1, a frequência mediana de 12, e a frequência média de 64.42 eventos por motorista.

FIGURA 17 – ESTATÍSTICAS X RECURSO



Fonte: O autor (2023)

O motorista com maior número de eventos acumulou 1.221 eventos, o que representa 7.1% do total de eventos registrados, seguido por outro motorista com 1.099 eventos, correspondendo a 6.39% do total. Esses dois motoristas combinados representam uma fração significativa do conjunto de dados, indicando uma concentração de eventos em poucos recursos. A frequência máxima registrada de 1.221 eventos é significativamente maior que a frequência média, o que sugere uma alta disparidade entre os motoristas em termos de participação nos eventos.

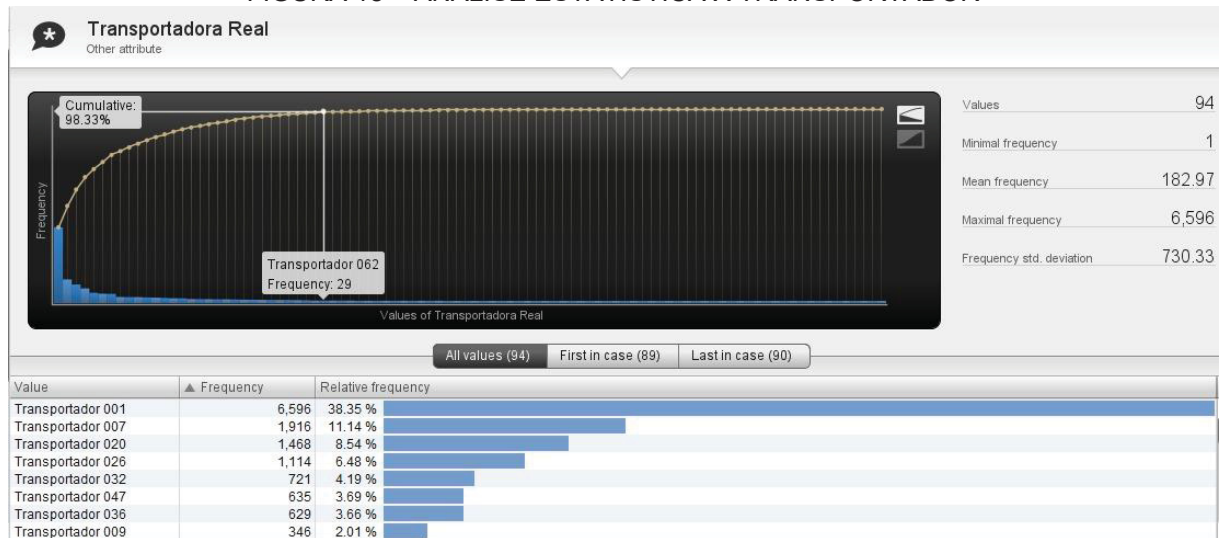
O desvio padrão de 145.12 reforça essa variabilidade, indicando que a maioria dos motoristas está associada a um número de eventos muito inferior ao dos motoristas com maiores frequências. Além disso, a tabela mostra que os motoristas com as maiores frequências de eventos representam uma parte considerável dos registros totais, enquanto muitos outros motoristas têm frequências significativamente menores.

Essa concentração de eventos em poucos motoristas pode ser indicativa de vários fatores, como maior envolvimento operacional desses indivíduos ou possíveis condições que favorecem a recorrência de eventos nesses casos específicos. No entanto, é importante investigar mais profundamente as razões dessa concentração, pois pode haver fatores externos, como o tipo de rota ou veículo, que estão contribuindo para essa discrepância nos dados.

A Figura 18 apresenta a distribuição de eventos associados a diferentes transportadoras no conjunto de dados. Ao todo, há 94 transportadoras registradas,

com uma frequência mínima de 1 evento e uma frequência máxima de 6.596 eventos para a transportadora com o maior número de registros.

FIGURA 18 – ANÁLISE ESTATÍSTICA X TRANSPORTADOR



Fonte: O Autor (2023)

A frequência média de registros por transportadora é de 182,97 eventos, e o desvio padrão de 730,33 reflete uma variação extremamente significativa entre as transportadoras, o que indica que algumas acumulam muitos eventos, enquanto a maioria tem uma presença muito menor.

A transportadora com maior número de registros, codificada como Transportadora 001, aparece com 6.596 registros, o que representa 38.35% do total de eventos. Isso demonstra uma concentração considerável de eventos em uma única transportadora. Em seguida, a Transportadora 007 aparece com 1.916 registros (11.14%), enquanto a Transportadora 020 tem 1.468 registros, correspondendo a 8.53% do total de eventos. As demais transportadoras apresentam frequências significativamente menores. Por exemplo, a Transportadora 026 tem 1.114 registros (6.48%), e a Transportadora 032 tem 721 registros (4.19%). A grande maioria das transportadoras registra uma quantidade muito menor de eventos, com várias delas tendo apenas alguns registros.

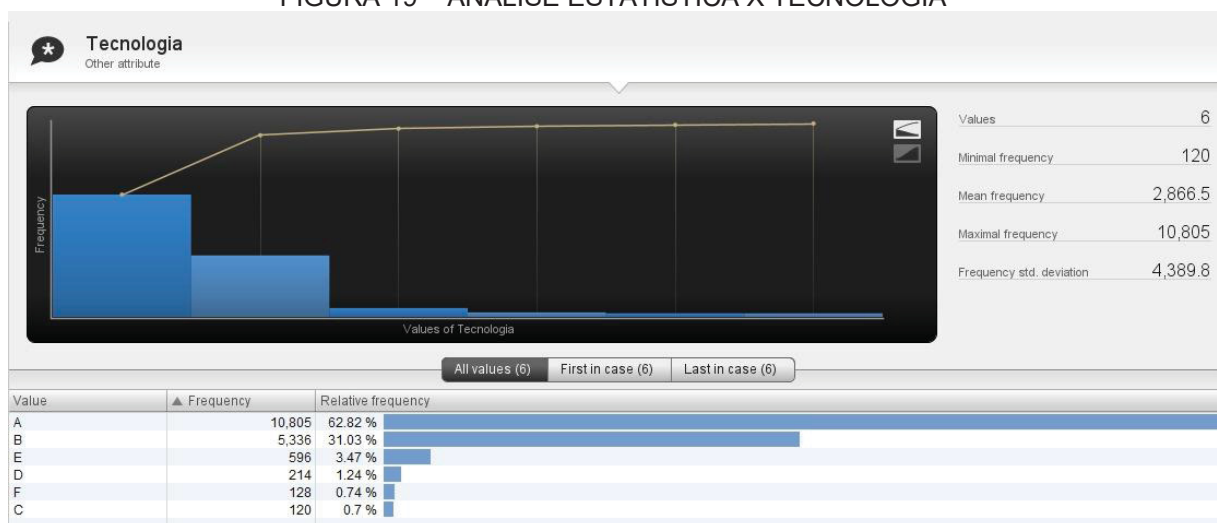
A disparidade entre a Transportadora 001 e as demais é evidente. A alta frequência registrada para essa transportadora pode indicar um papel central nas operações ou uma especialização que gera um volume elevado de eventos. O desvio padrão alto reflete a grande diferença entre as transportadoras mais e menos

frequentes, indicando uma distribuição desigual no número de eventos entre as diferentes transportadoras.

Essa concentração de eventos em poucas transportadoras, especialmente na Transportadora 001, sugere que essas empresas têm um papel muito mais relevante nas operações do que as outras. Pode ser o caso de uma maior participação de mercado ou operações específicas que geram mais eventos. Entretanto, o volume de registros requer uma análise cuidadosa para determinar as razões dessa concentração.

A Figura 19 apresenta a distribuição de eventos relacionados a diferentes tecnologias no conjunto de dados. No total, foram identificadas 6 tecnologias distintas, com uma frequência mínima de 120 eventos e uma frequência máxima de 10.805 eventos. A frequência média por tecnologia é de 2.866,5 eventos, enquanto o desvio padrão de 4.389,8 indica uma variação muito alta entre as frequências de uso das diferentes tecnologias.

FIGURA 19 – ANÁLISE ESTATÍSTICA X TECNOLOGIA



Fonte: O Autor (2023)

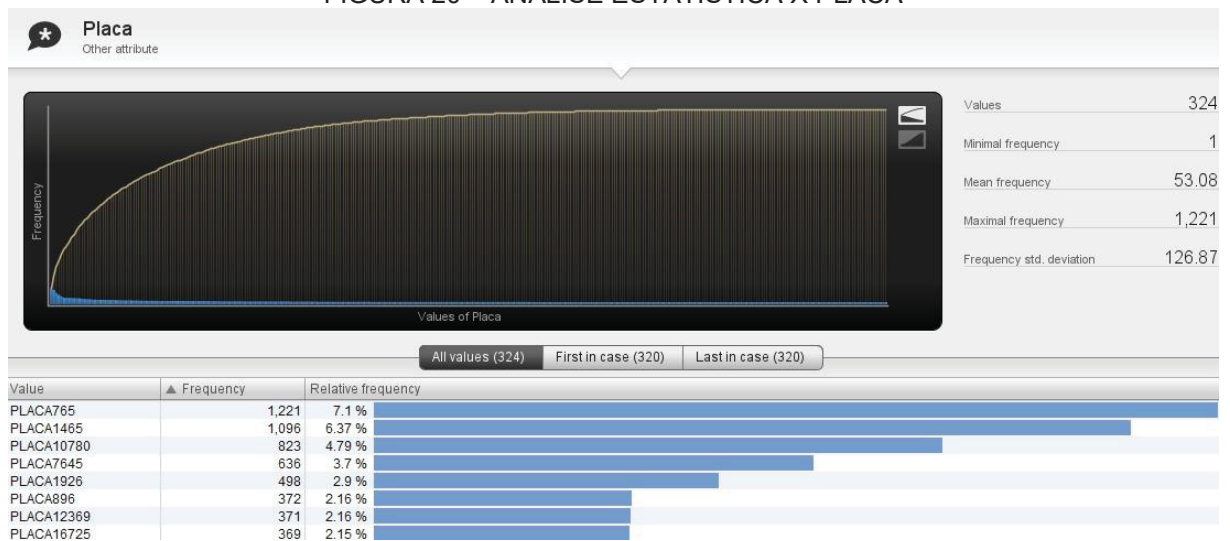
A Tecnologia A se destaca com 10.805 registros, o que representa 62.82% de todos os eventos. Isso evidencia o papel central dessa tecnologia nas operações registradas, sugerindo uma ampla adoção ou uma aplicação muito frequente. A Tecnologia B aparece em seguida com 5.336 registros, correspondendo a 31.03% do total, ainda substancial, mas consideravelmente menor que a Tecnologia A. As demais tecnologias têm uma participação muito menor. A Tecnologia E apresenta 596 registros (3.47%), a Tecnologia D aparece com 214 registros (1.24%), enquanto a

Tecnologia F tem 128 registros (0.74%) e a Tecnologia C tem a menor frequência, com 120 registros (0.7%).

O desvio padrão elevado reflete a distribuição desigual dos registros, com uma concentração muito forte em torno das Tecnologias A e B, que juntas somam mais de 90% dos eventos. Isso indica que, enquanto algumas tecnologias são amplamente utilizadas e registradas, as outras têm uma presença muito menor no conjunto de dados. A predominância de Tecnologia A e, em menor escala, Tecnologia B são reflexos da maior adoção pelos transportadores por terem sido as primeiras tecnologias homologadas na empresa conforme mapeado com especialista.

A Figura 20 apresenta a distribuição de eventos registrados por diferentes placas de veículos, com base nos motoristas envolvidos em acidentes sob sua responsabilidade. Ao todo, foram identificadas 324 placas únicas.

FIGURA 20 – ANÁLISE ESTATÍSTICA X PLACA



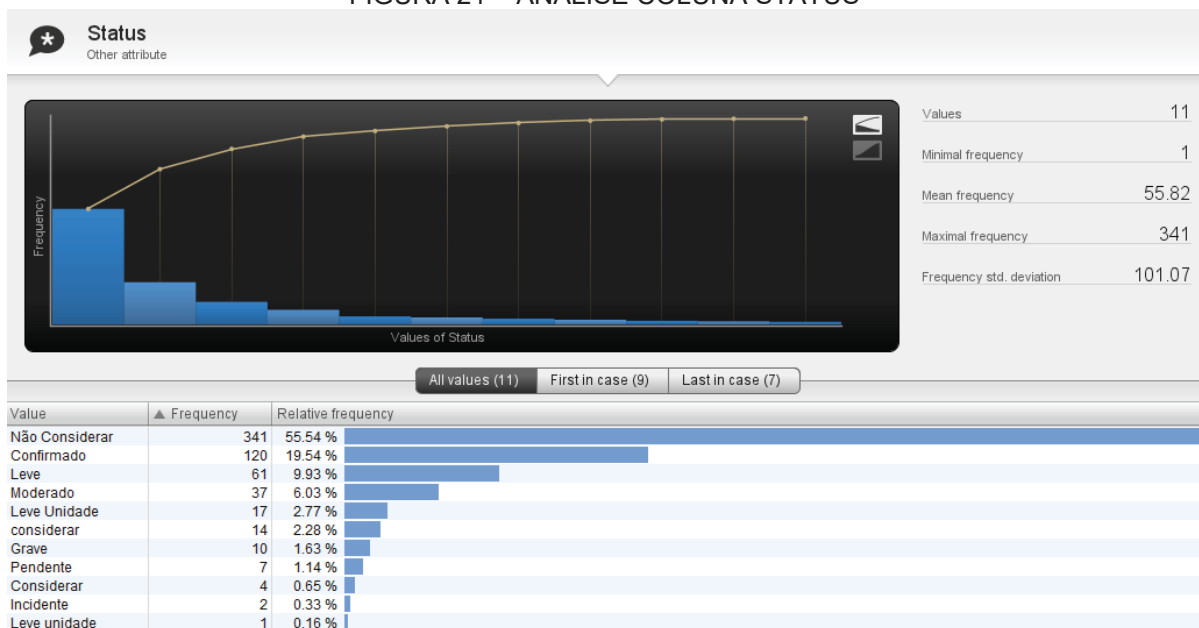
Fonte: O Autor (2023)

Conforme Figura 20, foram identificadas 324 placas únicas. A frequência mínima registrada para uma placa é 1 evento, enquanto a frequência média é de 53.08 eventos por placa. A frequência máxima é de 1.221 eventos, associada à placa com maior número de registros. O desvio padrão de 126.87 reflete uma variação significativa entre as placas, sugerindo que alguns veículos acumulam uma quantidade muito maior de eventos do que outros.

A placa mais frequente, PLACA765, aparece 1.221 vezes, o que representa 7.1% de todos os registros de eventos de responsabilidade do motorista. A segunda placa mais frequente, PLACA1465, acumula 1.096 registros (6.37%), seguida por

PLACA10780, com 823 registros (4.79%). Outras placas apresentam frequências progressivamente menores, como PLACA6745 com 636 registros (3.69%), e PLACA1926 com 498 registros (2.9%). O desvio padrão elevado indica que algumas placas estão envolvidas em uma proporção muito maior de eventos de acidentes. A próxima análise refere-se aos status informados na base de dados. Para cada alerta gerado é realizada uma análise pelo operador para avaliar considerar ou não no prontuário do motorista. Na Figura 21 é apresentada a síntese dos status gerados.

FIGURA 21 – ANÁLISE COLUNA STATUS



Fonte: O Autor (2023)

A coluna status é alimentada por um analista de SSMA que avalia os eventos mediante solicitação do transportador ou mediante auditoria. Conforme Figura 21 existem 11 valores de status diferentes registrados, O status menos frequente ocorre apenas uma vez. Em média, um status ocorre cerca de 55.82 vezes. O status mais frequente ocorre 341 vezes, o desvio padrão é 101.07, indicando uma variação significativa na frequência dos status.

De acordo com o gráfico o status Não Considerar é o status mais comum, ocorrendo 341 vezes, o que representa 55.54% do total de status registrados. Isso pode indicar que mais da metade dos eventos registrados são considerados não relevantes pós avaliação da equipe responsável. O segundo status mais comum é o confirmado, com 120 ocorrências, ou 19.54% do total. Esses podem ser casos que foram verificados e validados e ainda assim permaneceram na base. Os Status leve

e moderado apresentaram 61 ocorrências (9,93%) e 37 ocorrências (6,03%) estes referem-se à taxa de gravidade do acidente.

Os dados indicam que há uma concentração substancial de eventos classificados como "Não Considerar", merecem atenção principalmente por estarmos tratando de eventos em viagens com relato de acidentes. Desconsiderar os eventos de maneira errônea pode impactar negativamente na análise deixando passar algum comportamento importante.

#### 5.5.2.2 Mineração base de eventos sem acidentes

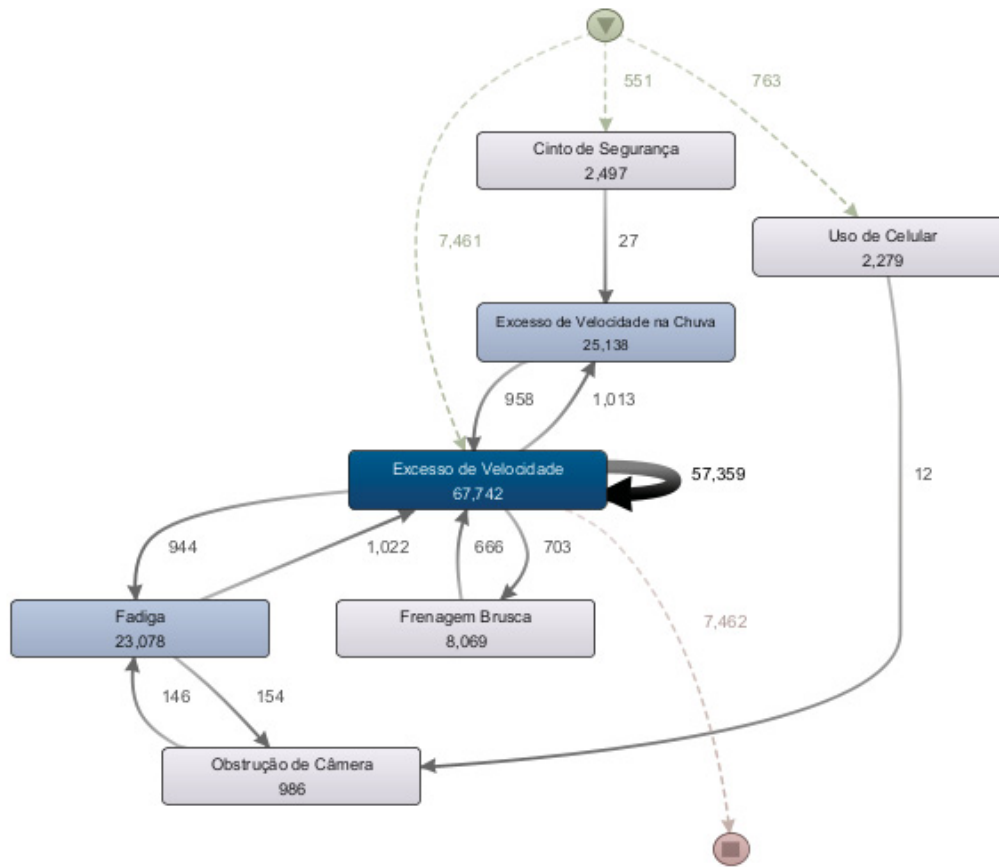
Nesta etapa da análise foram selecionados apenas eventos de viagens realizadas por motoristas que não tiveram acidentes no período avaliado.

Analisando a Figura 22, percebe-se que o "Excesso de Velocidade" aparece como o evento mais frequente, com um total de 67,742 ocorrências, o que sublinha a predominância desse comportamento de risco nas estradas. Interessante notar que, do total de eventos de excesso de velocidade, uma parcela significativa, exatamente 25,138 casos, ocorre sob condições climáticas adversas, como na chuva, evidenciando uma prática arriscada que agrava os perigos inerentes às estradas molhadas.

Além disso, a "Fadiga" contabiliza 23,078 incidências, sinalizando que o cansaço dos condutores representa um fator contribuinte notável para os riscos na direção. Da mesma forma, "Frenagem Brusca", registrada 8,069 vezes, pode indicar respostas súbitas a situações de perigo iminente ou até mesmo a uma distração no trânsito.

O alerta "Não utilização do Cinto de Segurança", com 2,497 registros, aponta para uma negligência preocupante com as medidas de segurança básicas. Enquanto isso, o "Uso de Celular" ao volante foi identificado em 2,279 ocasiões, reforçando a ideia de que a distração é um problema persistente e perigoso.

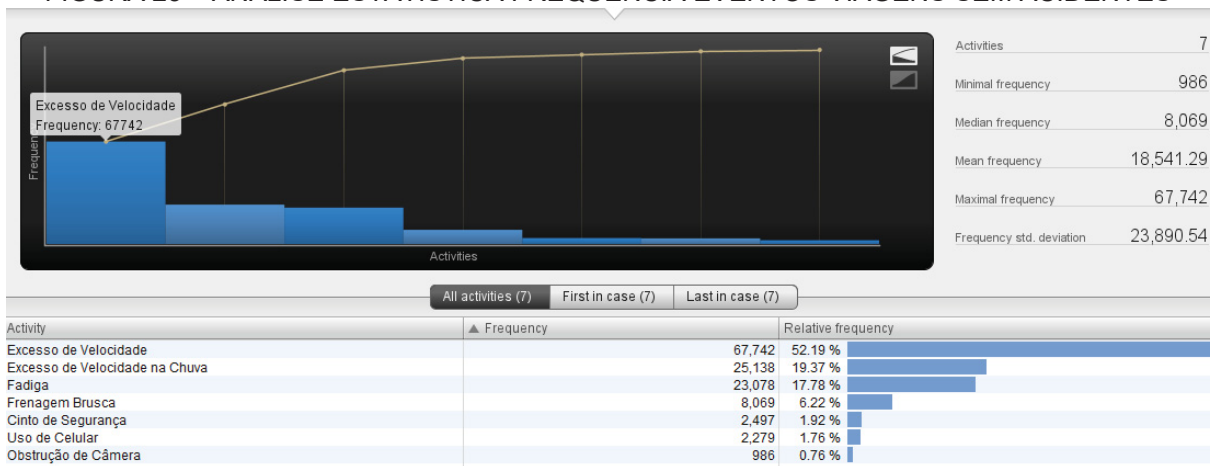
FIGURA 22 – FLUXO VIAGENS SEM ACIDENTES



Fonte: O Autor (2023)

A "Obstrução de Câmera" aparece em 986 eventos, sugerindo possíveis desafios no monitoramento adequado das condições de direção ou na manutenção do equipamento de segurança. A informação é vista de outra maneira na análise estatística conforme apresentado na Figura 23.

FIGURA 23 – ANÁLISE ESTATÍSTICA FREQUÊNCIA EVENTOS VIAGENS SEM ACIDENTES



Fonte: O Autor (2023)

Conforme Figura 23 eventos como "Frenagem Brusca" e "Uso de Celular" são menos frequentes, mas não menos significativos, correspondendo a 6,22% e 1,76%, respectivamente, do total de eventos. Estes indicam situações de reação a imprevistos e distração ao volante, que são fatores conhecidos por contribuir para acidentes de trânsito. Surpreendentemente, o "Cinto de Segurança" aparece em 1,92% dos casos, ressaltando um desafio contínuo na conformidade com as medidas básicas de segurança veicular.

O evento "Obstrução de Câmera", com apenas 0,76% das ocorrências, é o evento menos frequente. Isso pode sugerir que a obstrução de dispositivos de segurança é um problema menos prevalente ou que talvez seja subdetectado pelos sistemas de monitoramento atuais.

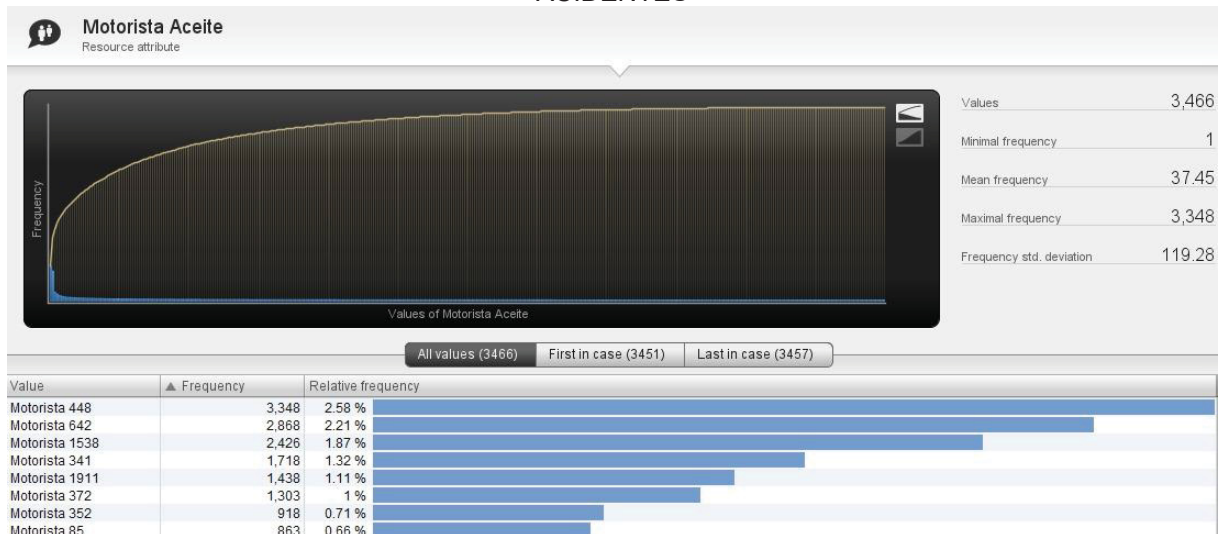
A frequência mediana dos eventos é de 8.069 ocorrências, sugerindo que metade dos eventos ocorre com mais frequência e a outra metade menos, apontando para uma distribuição desigual entre os tipos de eventos. Com uma média de frequência de 18.541,29 e um desvio padrão substancial de 23.890,54, há uma variação considerável na frequência de ocorrências entre os diferentes eventos.

O que chama a atenção é a variação significativa na frequência dos eventos entre os diferentes recursos. A frequência máxima é muito maior do que a média (36,49), o que indica que alguns recursos têm um número muito alto de eventos associados a eles, enquanto a maioria dos recursos tem um número de eventos mais próximo à mediana (10). O desvio padrão alto (117,84) reforça essa disparidade.

A Figura 24 apresenta a distribuição de frequência dos eventos categorizados pelos motoristas aceitos em viagens sem acidentes. O motorista com a maior frequência de eventos é o Motorista 448, com 3.348 ocorrências, o que representa 2,58% da frequência relativa total. Esse motorista está consideravelmente à frente dos demais em termos de participação nos eventos registrados.

Em segundo lugar, o Motorista 642 tem 2.868 eventos (2,21%), seguido pelo Motorista 1538, com 2.426 eventos (1,87%). Outros motoristas também se destacam, como o Motorista 341, com 1.719 eventos (1,32%), e o Motorista 1911, com 1.438 eventos (1,11%) .

FIGURA 24 – ANÁLISE ESTATÍSTICA FREQUÊNCIA EVENTOS x RECURSO VIAGENS SEM ACIDENTES



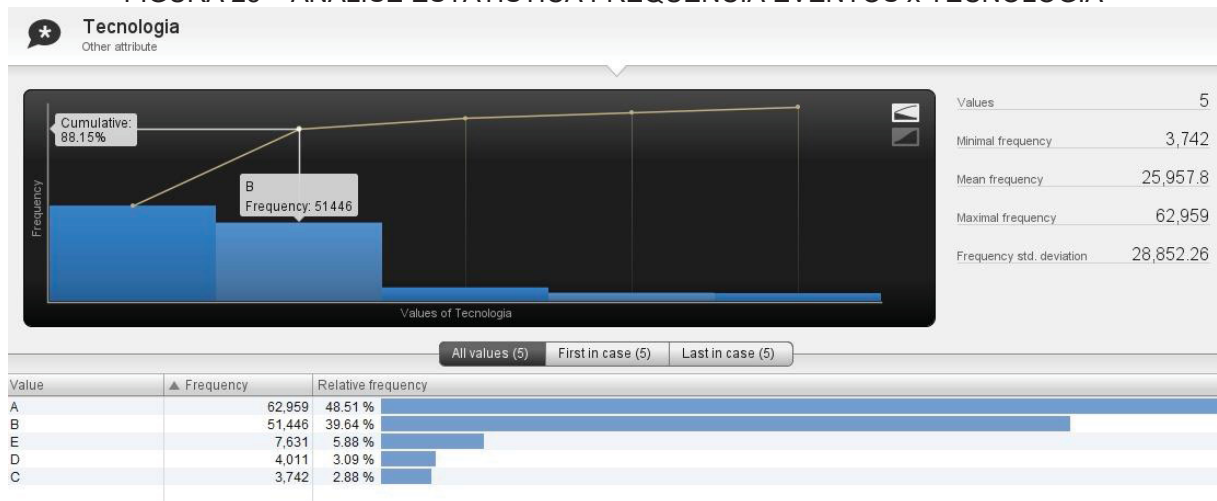
Fonte: O Autor (2023)

A base contém 3.466 motoristas únicos, com uma frequência mínima de 1 evento e uma frequência máxima de 3.348 eventos. A frequência média é de 37,45 eventos por motorista, e o desvio padrão de 119,28 sugere uma grande variação no número de eventos registrados por motorista, evidenciando que apenas alguns motoristas concentram a maior parte dos eventos.

A análise da distribuição de eventos por tecnologia utilizada em viagens sem acidentes, conforme demonstrado na Figura 25, revela uma clara predominância de dois fornecedores principais. A tecnologia A é a mais utilizada, com 62.959 eventos, representando 48,51% da frequência total. Essa concentração sugere uma ampla adoção dessa tecnologia pelos transportadores que não se envolveram em acidentes, o que pode estar relacionado à sua maior presença no mercado. A tecnologia B, com 51.446 eventos e 39,64% da frequência, também demonstra uma adoção significativa, reforçando o papel crucial dessas soluções no monitoramento e prevenção de incidentes.

As demais tecnologias, C, D e E, possuem frequências consideravelmente menores, com 3.742 eventos (2,88%), 4.011 eventos (3,09%) e 7.631 eventos (5,88%), respectivamente. Essa menor frequência pode refletir uma adoção mais limitada ou um uso específico por um número reduzido de transportadores.

FIGURA 25 – ANÁLISE ESTATÍSTICA FREQUÊNCIA EVENTOS x TECNOLOGIA



Fonte: O Autor (2023)

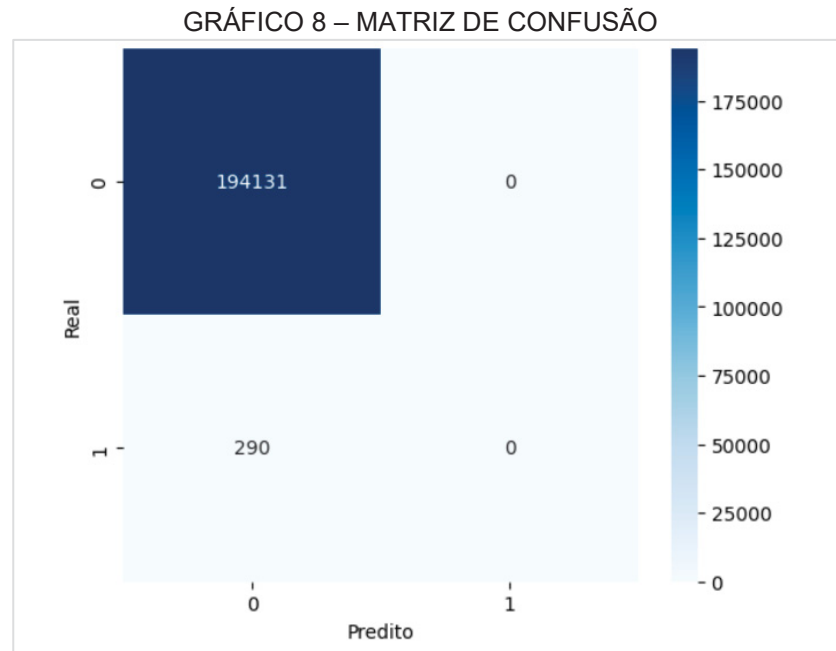
Os dados podem ser úteis para identificar recursos que estão frequentemente envolvidos em eventos, o que pode sugerir a necessidade de manutenção, treinamento ou outras intervenções. Além disso, o entendimento da distribuição de eventos por recurso pode ajudar na alocação eficiente de recursos e na priorização de ações dentro de uma estratégia de gestão operacional ou de riscos.

### 5.5.3 Análise de Correlação

Nesta etapa foi realizada uma análise de correlação entre os acidentes ocorridos e os eventos gerados pelas tecnologias. A avaliação foi realizada utilizando Análise de Regressão logística (Peng et.al.;2002). Considerando como variável resposta a coluna Acidente (0,1) e como variáveis preditoras os eventos, dia da semana e hora do dia, porém o modelo teve uma baixa acuracidade.

A matriz de confusão é uma ferramenta amplamente utilizada para avaliar o desempenho de modelos de classificação (Fawcett, 2006). Ela é composta por quatro quadrantes que comparam as classes preditas pelo modelo (0 ou 1) com as classes reais (0 ou 1). Esses quadrantes fornecem informações sobre o desempenho do modelo em termos de previsões corretas e incorretas para ambas as classes.

A seguir, são apresentados os resultados da análise, que utilizou a coluna "Acidente" como variável resposta, e as colunas "Evento" e "Tecnologia" como variáveis preditoras. Os resultados são demonstrados no Gráfico 8.



Verdadeiro Negativo (TN): 194.131 são os casos em que o modelo previu a classe 0 (negativa) e a classe real também era 0. O modelo teve um desempenho excelente ao prever a classe negativa corretamente.

Falso Positivo (FP): 0 indica que não houve casos em que o modelo previu a classe 1 (positiva) enquanto a classe real era 0. Isso é bom, pois mostra que o modelo não cometeu erros ao prever a classe positiva quando a verdadeira classe era negativa.

Falso Negativo (FN): 290 são os casos em que o modelo previu a classe 0 (negativa) mas a classe real era 1 (positiva). Estes são erros de tipo II, onde o modelo falhou em identificar a classe positiva corretamente.

Verdadeiro Positivo (TP): 0 indica que não houve casos em que o modelo previu corretamente a classe 1 (positiva) quando a classe real era de fato 1.

A partir desta matriz, podemos calcular métricas como precisão, recall (sensibilidade), especificidade e a pontuação F1. No entanto, a ausência de verdadeiros positivos (TP = 0) sugere que o modelo não conseguiu identificar corretamente nenhum dos casos positivos reais, o que é uma grande limitação, especialmente se a classe 1 for importante para prever (por exemplo, se a classe 1 for a presença de uma doença, o modelo não conseguiu identificar nenhum dos casos verdadeiros da doença).

Além disso, a precisão do modelo para a classe positiva não pode ser calculada (seria divisão por zero), e o recall para a classe positiva é 0, o que indica que o modelo é ineficaz para a classe de interesse. Isso pode ser um indicativo de que o modelo está enviesado para a classe mais frequente ou que há uma necessidade de rebalanceamento de classes, alterações no algoritmo ou nos parâmetros utilizados. Abaixo dados de acuracidade do modelo.

```
Accuracy: 0.9985  
Precision: 0.0000  
Recall: 0.0000  
F1-Score: 0.0000  
AUC-ROC: 0.7244
```

A *Accuracy* (Acurácia): 0.9985 indica que o modelo foi correto em 99.85% das vezes ao fazer uma previsão. No entanto, esta métrica pode ser enganosa se as classes estiverem desbalanceadas, o que parece ser o caso aqui, pois o modelo não identificou nenhum verdadeiro positivo.

*Precision* (Precisão): 0.0000 significa que o modelo não teve precisão na previsão de casos positivos, uma vez que não houve verdadeiros positivos (e, presumivelmente, poucos ou nenhum falso positivo). Isso sugere que, apesar de a acurácia ser alta, o modelo não é útil para prever a classe de interesse (classe 1).

*Recall* (Revocação ou Sensibilidade): 0.0000 indica que o modelo não foi capaz de identificar corretamente nenhum dos casos positivos reais. Todos os casos positivos foram perdidos pelo modelo (290 falsos negativos).

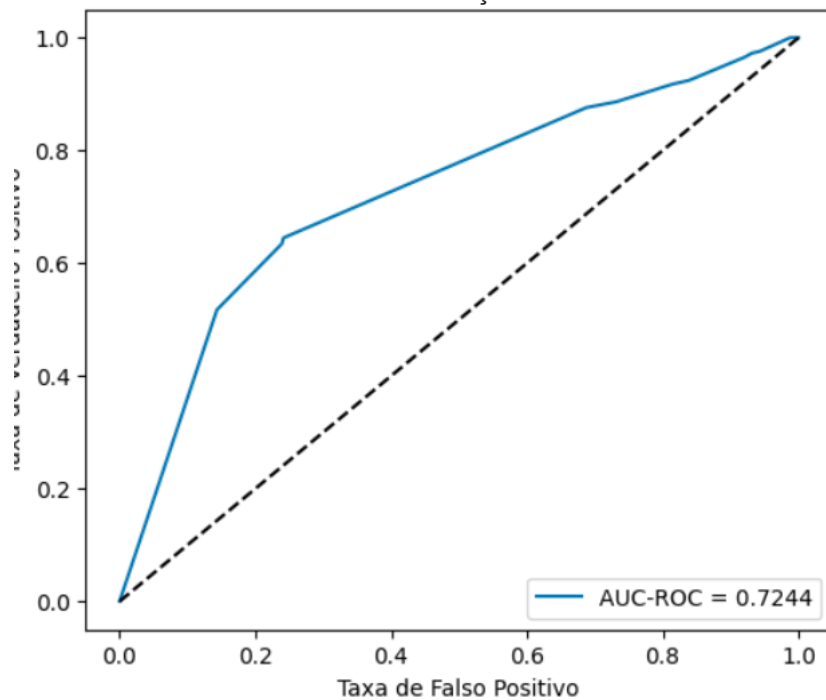
F1-Score: 0.0000 é a média harmônica de precisão e recall e neste caso também é zero, sugerindo que o modelo tem um desempenho muito pobre na classe de interesse. É especialmente importante quando se quer um equilíbrio entre precisão e recall e, neste caso, indica que o modelo é ineficaz.

AUC-ROC: 0.7244 refere-se à Área Sob a Curva da Característica de Operação do Receptor (Area Under the Receiver Operating Characteristic Curve). Uma AUC de 0.7244 é considerada aceitável e sugere que o modelo tem uma capacidade razoável de distinguir entre as classes positiva e negativa. No entanto, a utilidade desta métrica é limitada neste contexto, pois sabemos que o modelo falha em identificar verdadeiros positivos.

Dado o alto valor de acurácia e a AUC-ROC (Gráfico 9) relativamente boas, mas com precisão, recall e F1-Score sendo 0, é provável que o modelo esteja sofrendo

com um desequilíbrio de classes severo. Isso significa que a classe negativa é muito mais prevalente do que a classe positiva, e o modelo optou por prever quase sempre a classe negativa, ignorando a classe positiva, que é de maior interesse. Conforme explicado por Datacamp (2023), em casos de desequilíbrio de classes, a acurácia pode ser enganosa, pois o modelo foca na classe majoritária, o que resulta em previsões inadequadas para a classe minoritária. Além disso, Neptune.ai (2023) sugere que, embora o valor de AUC-ROC seja útil para avaliar o desempenho de modelos, ele pode não ser suficiente para lidar com o desbalanceamento de classes. Nesse sentido, o uso de métricas como F1-Score e Precision-Recall AUC é recomendado para avaliar corretamente o desempenho nas classes minoritárias, como em cenários de detecção de fraudes ou falhas (PLOS DIGITAL HEALTH, 2023).

GRÁFICO 9 – CURVA ROC PARA AVALIAÇÃO DA ACURACIDADE DO MODELO



Fonte: O Autor (2024)

Portanto, mesmo que o modelo pareça ter um desempenho excelente em termos de acurácia, na realidade, ele não está conseguindo capturar a complexidade dos dados e não é útil para previsões práticas.

## 5.6 AVALIAÇÃO

Ao longo do estudo, um dos desafios mais críticos foi relacionado à qualidade e à heterogeneidade dos dados. As bases de dados utilizadas eram provenientes de diferentes fontes e formatos, como planilhas Excel e bancos de dados Access, além de sistemas de monitoramento que geravam eventos variados, como telemetria, fadiga e acidentes. Embora existam regras estabelecidas para ajustar os dados, o fato de esses ajustes serem realizados de forma manual apresenta riscos significativos, especialmente no que tange à consistência e precisão dos dados.

Foi identificado que diferentes usuários, em diversas fases da coleta e processamento dos dados, realizaram ajustes manuais para corrigir falsos positivos ou categorizar eventos de maneira específica. Embora essas ações sigam procedimentos, a natureza manual dos ajustes cria potenciais inconsistências. Por exemplo, alguns eventos foram categorizados como "não considerar", representando 55,4% dos registros relacionados a acidentes. Essa categorização, se realizada sem uma padronização robusta, pode impactar a representatividade da amostra e a qualidade da análise final. Esse desafio destaca a importância de melhorar o processo de coleta e tratamento de dados, reforçando a necessidade de automatizar etapas críticas para minimizar o risco de erros e garantir que os ajustes sejam documentados de maneira clara e auditável.

As análises realizadas, utilizando técnicas de mineração de processos e análise de correlação, revelaram achados importantes sobre os comportamentos de risco dos motoristas e as condições que precedem os acidentes. Aqui estão os principais resultados de cada etapa:

A análise dos dados de acidentes revelou uma redução de mais de 50% nos acidentes moderados ao longo do período estudado. Essa redução, no entanto, não pode ser atribuída a novas intervenções realizadas durante o período, mas sim ao trabalho contínuo de melhoria das operações de segurança, que já vinha sendo implementado anteriormente. As práticas e políticas de segurança que foram mantidas ao longo do tempo parecem ter sido eficazes em reduzir a severidade de alguns acidentes, especialmente os de gravidade moderada.

Por outro lado, observou-se um aumento significativo nos acidentes leves, o que gera uma preocupação. Embora os acidentes leves não apresentem riscos imediatos à integridade física grave, eles representam uma ameaça latente, pois

podem evoluir para acidentes mais graves se não forem monitorados e tratados adequadamente. Essa tendência destaca a necessidade de continuar investindo em estratégias preventivas, mesmo diante da redução dos acidentes mais graves.

Em relação aos comportamentos de risco, o excesso de velocidade foi o fator mais prevalente, aparecendo em 92% dos acidentes analisados. Esse dado reforça a importância de manter um controle rigoroso sobre a condução em alta velocidade, especialmente em condições adversas. A fadiga, presente em 38% dos acidentes, também foi um fator crítico identificado, indicando a necessidade de revisar e fortalecer as políticas de controle de jornada e descanso dos motoristas. Esse fator destaca a urgência em adotar práticas que promovam o bem-estar dos motoristas, reduzindo a probabilidade de incidentes relacionados à fadiga.

A mineração de processos permitiu identificar padrões comportamentais que levam à ocorrência de acidentes. O evento mais comum foi o excesso de velocidade, seguido pela fadiga. A análise também revelou que eventos como frenagem brusca e uso de celular, embora menos frequentes, estão correlacionados com acidentes, principalmente aqueles de maior gravidade. Além disso, observou-se que, em condições adversas como a chuva, o excesso de velocidade está diretamente relacionado ao aumento do risco de acidentes graves.

Para facilitar a visualização foi elaborada uma tabela comparativa entre os dados das viagens de motoristas com acidentes x eventos de viagens de motoristas sem acidentes conforme apresentado na Tabela 2.

TABELA 3 – RESUMO RESULTADOS

Alerta	Base com Acidentes		Base sem Acidentes		Diferença	
	QTD	%	QTD	%	QTD	%
Excesso de Velocidade	5.534	92%	67.142	86%	61.608	6%
Excesso de Velocidade na Chuva	1.931	31%	25.138	38%	23.207	-7%
Frenagem Brusca	1.387	25%	8.069	12%	6.682	13%
Uso de Celular	120	2%	2.279	3%	2.159	-1%
Obstrução de Câmera	306	6%	986	1%	680	4%
Cinto de Segurança	1.572	28%	2.497	4%	925	25%
Fadiga	6.221	38%	23.078	35%	16.857	4%

Fonte: O Autor (2023)

A comparação realizada na Tabela 2 revela que os comportamentos de risco, como excesso de velocidade, frenagem brusca e fadiga, são significativamente mais

prevalentes entre os motoristas que se envolveram em acidentes. A diferença de 25% no uso do cinto de segurança sugere que a falta de adesão às normas de segurança contribui para a gravidade dos acidentes, reforçando a importância de monitorar esse comportamento de forma mais ativa.

A análise comparativa dos eventos críticos mostrou que motoristas envolvidos em acidentes tendem a exibir comportamentos de risco em proporções muito maiores do que aqueles que não se envolveram em acidentes. O excesso de velocidade, presente em 92% dos casos de motoristas com acidentes, se destaca como o principal fator de risco, o que está alinhado com a literatura que demonstra a correlação entre velocidade excessiva e acidentes graves (Shinar, 2007).

Além disso, eventos como frenagem brusca e obstrução de câmera também foram mais prevalentes entre motoristas com acidentes, o que pode indicar uma perda de controle do veículo ou falta de visibilidade adequada. Já o uso de celular foi mais prevalente entre motoristas sem acidentes, sugerindo que, embora seja um comportamento arriscado, não resultou diretamente em um aumento imediato de acidentes nesta base de dados específica.

Outro ponto importante identificado foi a diferença no uso do cinto de segurança. Entre os motoristas envolvidos em acidentes, 28% não estavam utilizando o cinto, comparado a apenas 4% entre os motoristas sem acidentes. Esse dado reforça a necessidade de campanhas e monitoramento para garantir a adesão ao uso de dispositivos de segurança, especialmente para reduzir a gravidade dos acidentes.

A análise de correlação entre os eventos e os acidentes utilizando regressão logística apresentou limitações devido ao desbalanceamento de classes. A acurácia do modelo foi comprometida, sugerindo que o conjunto de dados necessita de técnicas mais avançadas de rebalanceamento ou a inclusão de novas variáveis preditoras para melhorar a capacidade do modelo de prever acidentes. Os eventos de excesso de velocidade e fadiga novamente se destacaram como os principais preditores de acidentes, reforçando a validade dos achados anteriores.

Após a realização das análises, foi necessário validar os resultados com as partes interessadas e garantir que os achados estivessem de acordo com as expectativas e as melhores práticas da indústria. As entrevistas com especialistas em segurança no transporte confirmaram que os comportamentos de excesso de velocidade e fadiga são de fato precursores críticos de acidentes, o que valida a

análise realizada. No entanto, também foram destacadas áreas que precisam de maior atenção, como o controle dos ajustes manuais realizados nos dados.

Os especialistas sugeriram que, embora existam regras para ajustar os dados e categorizar eventos como "não considerar", a natureza manual desses ajustes traz o risco de erros ou inconsistências. Para mitigar esses riscos, foi recomendada a implementação de ferramentas que automatizem parte desse processo e garantam que os ajustes sejam feitos de forma padronizada e auditável. Isso não apenas aumentaria a qualidade dos dados, mas também fortaleceria a confiança nos resultados das análises.

Durante a validação quantitativa, a comparação com benchmarks internos decorrentes das análises mensais realizadas pela equipe de segurança do transporte mostrou que a taxa de excesso de velocidade está dentro dos padrões esperados. No entanto, a taxa de não uso do cinto de segurança foi consideravelmente maior do que o esperado, indicando que essa é uma área que merece maior atenção dentro da operação analisada.

## 5.7 MELHORIAS E SUPORTE

Com base nos resultados obtidos nas análises de acidentes e na mineração de processos, foi possível identificar áreas críticas que demandam melhorias imediatas e suporte contínuo. Embora 90% da frota já esteja equipada com **sistemas de telemetria** e 75% com **sistemas de monitoramento de fadiga**, os dados indicam que ainda há comportamentos de risco recorrentes que precisam ser abordados por meio de ações corretivas e melhorias no processo de gestão.

### 5.7.1 Ações Propostas

1. Aprimoramento do Monitoramento de Comportamentos de Risco Apesar da alta cobertura de telemetria, a análise revelou que o excesso de velocidade permanece como o fator mais prevalente, presente em 92% dos acidentes. Estudos mostram que monitorar o comportamento dos motoristas em tempo real e intervir preventivamente pode reduzir significativamente os acidentes, principalmente com sistemas de telemetria avançada (CAIRD et al., 2014;

DONMEZ, BOYLE e LEE, 2008). Assim, propõe-se o aprimoramento dos alertas de telemetria, com foco em situações críticas, como excesso de velocidade em condições adversas, além da implementação de notificações em tempo real para intervenções imediatas.

2. **Expansão e Otimização dos Sistemas de Monitoramento de Fadiga** A fadiga foi associada a 38% dos acidentes, destacando a necessidade de uma gestão mais rigorosa das jornadas de trabalho dos motoristas. A literatura indica que sistemas de monitoramento contínuo da fadiga, utilizando tecnologias como câmeras infravermelhas, têm sido eficazes na detecção precoce da sonolência e podem reduzir significativamente o número de acidentes (MAY e BALDWIN, 2009; SAYED e ESKANDARIAN, 2020). Assim, recomenda-se a expansão dos sistemas de monitoramento de fadiga para toda a frota e a otimização dos alertas gerados, garantindo que os riscos sejam identificados e mitigados.
3. **Reforço no Uso de Cinto de Segurança** A adesão ao uso do cinto de segurança é uma questão crítica, e estudos indicam que combinar o uso de telemetria com campanhas de conscientização pode aumentar significativamente a adesão dos motoristas (WHO, 2018; ROEHLER, SLEET e NAUMANN, 2013). A telemetria deve incluir mecanismos de alerta para o uso do cinto de segurança, além de campanhas contínuas de conscientização para garantir que os motoristas adotem essa prática de segurança básica.
4. **Automatização do Monitoramento e Relatórios** A automação da coleta de dados e dos relatórios, utilizando sistemas integrados, pode reduzir erros manuais e aumentar a eficácia do monitoramento, como apontado em estudos de grandes volumes de dados provenientes de telemetria (LAURILA e PAJUNEN, 2014; GHILARDI et al., 2016). Propõe-se a automatização da coleta de dados, eliminando a necessidade de intervenções manuais e aumentando a confiabilidade dos relatórios gerados.

### 5.7.2 INDICADORES PARA ACOMPANHAMENTO

Os KPIs indicadores sugeridos para medir o sucesso das ações propostas incluem:

- Redução dos eventos de excesso de velocidade em 10% nos próximos 6 meses.
- Aumento da cobertura do sistema de fadiga para 100% da frota no próximo ano.
- Diminuição de acidentes moderados e graves em 20% ao longo de 12 meses.
- Adesão ao uso de cinto de segurança superior a 95% dos motoristas.

### 5.7.3 MONITORAMENTO CONTÍNUO E SUPORTE OPERACIONAL

As ações sugeridas deverão ser acompanhadas de um monitoramento contínuo, com revisões regulares dos resultados dos indicadores e feedback direto dos motoristas e gestores sobre as novas práticas implementadas. Relatórios periódicos gerados pela telemetria serão utilizados para ajustar as intervenções, garantindo que os comportamentos de risco sejam minimizados ao longo do tempo.

## 6 APRESENTAÇÃO DOS RESULTADOS

Nesta seção, são apresentados os principais resultados obtidos a partir da revisão da literatura, do desenvolvimento do framework e da aplicação da mineração de processos no contexto de gestão de acidentes rodoviários.

A revisão da literatura foi conduzida utilizando o método PRISMA(MOHER et al., 2010) com o objetivo de identificar os principais estudos sobre as causas de acidentes rodoviários e as técnicas aplicadas para sua redução. Além disso, foram revisados artigos que aplicaram a mineração de processos (MP) no contexto do transporte. Os principais resultados dessa revisão incluem:

- A maioria dos estudos foca em comportamentos de risco, como excesso de velocidade, fadiga do motorista, e uso de celular, como fatores contribuintes para acidentes rodoviários.
- Foram encontrados vários artigos que utilizam MP em setores de transporte e logística para otimização de rotas, controle de frota, e prevenção de falhas mecânicas.
- No entanto, não foram encontrados estudos que aplicassem diretamente a MP no contexto de gestão de acidentes rodoviários, evidenciando uma lacuna importante na literatura.
- Essa revisão reforça a necessidade de um framework específico para aplicação de MP na gestão de acidentes, com foco na identificação de padrões comportamentais de risco que possam ser mitigados.

Com base na lacuna identificada, foi desenvolvido um framework utilizando a metodologia Design Science Research (DRESCH; LACERDA; ANTUNES JÚNIOR, 2015). O framework é composto por sete etapas principais, projetadas para aplicar a mineração de processos à gestão de acidentes rodoviários:

- Coleta de Dados: A primeira etapa envolveu a definição dos tipos de dados relevantes (telemetria, registros de acidentes, monitoramento de fadiga).
- Pré-processamento dos Dados: Os dados foram limpos e padronizados, eliminando inconsistências e duplicidades.
- Mineração de Processos: Aplicação de algoritmos de MP para identificar padrões de risco nos dados.

- **Análise de Padrões:** Identificação de padrões de comportamento críticos que precedem os acidentes.
- **Análise Comparativa:** Comparação dos eventos entre motoristas que sofreram acidentes e aqueles que não sofreram.
- **Correlação dos Dados:** Análise de correlação entre eventos e acidentes, permitindo a identificação de comportamentos de risco mais prováveis.
- **Geração de Relatórios e Recomendações:** A partir dos padrões identificados, foram feitas recomendações para intervenções preventivas.

O framework foi desenhado para ser aplicado em bases de dados de grandes volumes, como no setor de transporte de cargas, e sua flexibilidade permite ser adaptado para diferentes tipos de acidentes e comportamentos de risco. O framework foi aplicado em um estudo de caso com dados reais no transporte de longo curso de uma multinacional do ramo alimentício. A base de dados utilizada continha 17.199 registros relacionados a eventos de telemetria e acidentes ao longo de um período de 12 meses. Os resultados desta aplicação foram os seguintes:

1. A aplicação do framework permitiu identificar que excesso de velocidade e fadiga eram os principais comportamentos de risco relacionados aos acidentes. O excesso de velocidade foi detectado em 92% dos casos, e a fadiga foi associada a 38% dos acidentes.
2. Quando comparados os motoristas que sofreram acidentes com aqueles que não sofreram, observou-se que os primeiros apresentaram uma maior incidência de comportamentos de risco, como excesso de velocidade e uso de celular.
3. A MP permitiu mapear a sequência de eventos que culminavam em acidentes, revelando padrões consistentes, como frenagens bruscas precedidas por excesso de velocidade em condições adversas.

Os resultados da mineração de processos mostraram padrões claros de comportamento de risco, permitindo que se fizessem as seguintes conclusões:

**Correlação com Tecnologia:** Tecnologias como sistemas de telemetria e monitoramento de fadiga desempenharam um papel crucial na detecção e prevenção de comportamentos de risco, mas a análise de correlação sugere que esses sistemas precisam ser aprimorados para detectar padrões mais complexos.

A análise comparativa das bases de motoristas com e sem acidentes revelou importantes diferenças nos eventos de segurança:

- Esses dados confirmam que o excesso de velocidade e a fadiga são os principais fatores de risco que diferenciam motoristas que se envolvem em acidentes daqueles que não se envolvem.
- Os resultados obtidos com a aplicação do framework indicam que a mineração de processos é uma ferramenta valiosa para a gestão de acidentes rodoviários. Os padrões identificados fornecem uma base para intervenções que podem ser implementadas no futuro, como políticas de controle de jornada, campanhas de conscientização sobre o uso de celular e treinamentos focados em reduzir o excesso de velocidade.
- O framework desenvolvido mostrou-se eficiente na detecção de padrões complexos de comportamento, e a sua aplicação pode ser expandida para outros setores ou áreas de segurança no transporte, oferecendo uma abordagem proativa para a prevenção de acidentes.

## 7 CONCLUSÃO

Este estudo teve como objetivo central o desenvolvimento e aplicação de um framework, o PM2RCS, para a gestão de acidentes rodoviários no transporte de cargas, utilizando a mineração de processos como base analítica. Através da análise de dados históricos de telemetria, sensores de fadiga e registros de acidentes, o framework foi testado em um contexto operacional real, demonstrando sua eficácia na identificação de padrões críticos de risco e comportamentos que contribuem para acidentes rodoviários.

O estudo abordou, em primeiro lugar, uma revisão da literatura, conduzida com o método PRISMA, que destacou uma lacuna significativa na aplicação da mineração de processos no gerenciamento de acidentes. A análise da literatura revelou que, apesar de a mineração de processos ser amplamente aplicada em setores como transporte e logística para otimização de operações e rotas, sua utilização voltada para a gestão de acidentes rodoviários não havia sido suficientemente explorada. Isso reforçou a relevância e necessidade de um framework dedicado à identificação de padrões que levem à prevenção de acidentes, especificamente no contexto rodoviário, onde as variáveis são complexas e frequentemente dependem de comportamentos humanos.

O desenvolvimento e a estruturação da base de dados representaram um segundo ponto crucial no alcance dos objetivos do estudo. Durante essa fase, foram enfrentados desafios relacionados à qualidade e padronização dos dados, uma vez que os registros eram provenientes de fontes diversas, com formatos variados e níveis distintos de granularidade. A integração de dados de diferentes tecnologias, como sistemas de telemetria e monitoramento de fadiga, exigiu um esforço considerável de pré-processamento para garantir a consistência das informações. Esse processo foi essencial para garantir a integridade dos dados, que sustentariam as análises realizadas na etapa de mineração de processos.

A aplicação das técnicas de mineração de processos, objetivo central do framework, revelou insights profundos sobre os comportamentos de risco mais frequentes. O excesso de velocidade destacou-se como o evento mais prevalente, identificado em 92% dos acidentes analisados, o que reforça a necessidade de políticas de controle rigoroso de velocidade em operações rodoviárias. Além disso, a fadiga dos motoristas foi outro fator crítico, presente em 38% dos casos, evidenciando

a importância de iniciativas que promovam o descanso adequado e o cumprimento das normas de jornada de trabalho. Esses resultados indicam que a mineração de processos tem potencial para detectar padrões recorrentes e, conseqüentemente, atuar de forma preventiva na mitigação de comportamentos que elevam o risco de acidentes.

A eficácia do framework PM2RCS também foi avaliada em termos de sua capacidade de gerar recomendações práticas para a melhoria da gestão de segurança no transporte rodoviário. A aplicação prática no estudo de caso revelou que, ao analisar grandes volumes de dados com precisão e eficiência, o framework permite que gestores identifiquem áreas críticas e intervenham de maneira mais direcionada. A telemetria e os sensores de fadiga foram componentes-chave para a coleta de dados, oferecendo uma visão contínua das condições operacionais, algo que potencializa a capacidade de reação rápida em situações de risco.

Do ponto de vista teórico, o estudo oferece uma contribuição significativa ao integrar a mineração de processos em um contexto onde sua aplicação ainda era limitada. O framework proposto não só demonstrou sua viabilidade técnica, como também ampliou o escopo de uso da MP em contextos de segurança no transporte, proporcionando uma nova perspectiva para o uso dessa técnica em análises preventivas. Metodologicamente, a adoção da Design Science Research (DSR) provou ser uma abordagem eficaz para o desenvolvimento, implementação e avaliação de soluções tecnológicas para problemas complexos, como a gestão de acidentes rodoviários.

No entanto, o estudo também enfrentou desafios. A falta de padronização nas bases de dados e a manipulação manual de informações introduziram limitações que, em parte, comprometeram a consistência dos resultados obtidos. Embora as tecnologias de monitoramento de telemetria e fadiga estejam amplamente implementadas, cobrindo 90% e 75% da frota, respectivamente, a qualidade dos dados coletados ainda pode ser aprimorada. Isso ressalta a importância de investimentos contínuos em governança de dados e na adoção de tecnologias mais integradas, que possam automatizar e padronizar a coleta e análise de informações, minimizando o risco de erros e aumentando a eficácia do framework.

Outro ponto que merece atenção é a resistência cultural e organizacional à adoção de novas tecnologias e metodologias analíticas no setor de transporte rodoviário de cargas. Esse fator limitou, em parte, a velocidade de implementação de

medidas preventivas mais sofisticadas e tecnológicas. Para superar essa barreira, é necessária uma abordagem que combine a aplicação de soluções tecnológicas com iniciativas de capacitação e conscientização dos profissionais da área.

Concluindo, o framework PM2-RCS demonstrou ser uma ferramenta valiosa para a gestão de segurança no transporte rodoviário, possibilitando a identificação de comportamentos de risco e fornecendo uma base sólida para a implementação de políticas preventivas. Os resultados obtidos indicam que a mineração de processos é capaz de fornecer insights profundos e de relevância prática, contribuindo para a melhoria da segurança operacional e a redução de acidentes. A aplicação bem-sucedida do framework abre novas oportunidades para pesquisas futuras, especialmente no que se refere à integração de tecnologias emergentes, como inteligência artificial e aprendizado de máquina, que podem aprimorar a análise e a previsão de riscos de forma ainda mais eficiente. Assim, este estudo oferece uma base sólida para o desenvolvimento de práticas operacionais mais seguras e sustentáveis, atendendo tanto às necessidades do setor quanto às demandas de segurança pública.

## REFERÊNCIAS

- AHMED, A.; SADULLAH, A. F. M.; YAHYA, A. S. Errors in accident data, its types, causes and methods of rectification-analysis of the literature. **Accident Analysis and Prevention**, v. 130, p. 3–21, 1 set. 2019.
- AIDMAN, E. et al. Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. **Accident Analysis and Prevention**, v. 81, p. 8–13, 1 ago. 2015.
- AIT-MLOUK, A.; GHARNATI, F.; AGOUTI, T. An improved approach for association rule mining using a multi-criteria decision support system: a case study in road safety. **European Transport Research Review**, v. 9, n. 3, 1 set. 2017.
- BAIKEJULI, M.; SHI, J.; HUSSAIN, M. A study on the probabilistic quantification of heavy-truck crash risk under the influence of multi-factors. **Accident Analysis and Prevention**, v. 174, 1 set. 2022.
- BARUA, D.; JAIN, P.; GUPTA, J.; GADRE, D. V. Road Accident Prevention Unit (R.A.P.U): A prototyping approach to mitigate an omnipresent threat. 2013. Disponível em: <https://doi.org/10.48550/arXiv.1311.1089>. Acesso em: 23 out. 2022.
- BAYAZIT, N. Investigating Design: A Review of Forty Years of Design Research. **Design Issues**, v. 20, n. 1, p. 16–29, jan. 2004.
- CATALANO, M.; GALATIOTO, F.; SHAIKH, N. Accident-related cost analysis and decision-making support through econometric modelling. 2020.
- CHINTALAPATI, S. S.; KRISHNA, C.; MADHAV, B. T. P. Analysis on IoT Environment for Detection and Prevention of Road Accidents with Communication Modules. **Journal of Green Engineering (JGE)**. [S.l.: s.n.].
- CNT - Confederação Nacional dos Transportes. Disponível em: <https://www.cnt.org.br/painel-acidente>. Acesso em: 2 ago. 2022.
- DE OÑA, J.; LÓPEZ, G.; ABELLÁN, J. Extracting decision rules from police accident reports through decision trees. **Accident Analysis and Prevention**, v. 50, p. 1151–1160, jan. 2013.
- DELORME, R.; LASSARRE, S. A new theory of complexity for safety research. The case of the long-lasting gap in road safety outcomes between France and Great Britain. **Safety Science**, v. 70, p. 488–503, 2014.
- DINGUS, T. A.; GUO, F.; LEE, S.; ANTIN, J. F.; PEREZ, M. A.; BERMUDEZ, M.; HANKS, D. C. Driver crash risk factors and prevalence evaluation using naturalistic driving data. **Proceedings of the National Academy of Sciences**, v. 113, n. 10, p. 2636-2641, 2016.

- DOROFEEV, A. et al. Ensuring the integrity of transportation and logistics during the COVID-19 pandemic. **Transportation Research Procedia**. Anais... Elsevier B.V., 2020.
- DRAMSKI, M. Extensible event stream format for navigational data. **Scientific Journals of the Maritime University of Szczecin**, v. 47, 2016.
- DRAMSKI, M. Guidelines for recording transport event logs. **Scientific Journals of the Maritime University of Szczecin**, v. 53, n. 125, p. 102–105, 2018a.
- DRAMSKI, M. Guidelines for recording transport event logs. **Scientific Journals of the Maritime University of Szczecin**, v. 53, n. 125, p. 102–105, 2018b.
- DRESCH, Aline; LACERDA, Daniel Pacheco; ANTUNES JÚNIOR, José Antônio Valle. *Design Science Research: A Method for Science and Technology Advancement*. **Springer**, 2015.
- ELVIK, R. Risk of road accident associated with the use of drugs: A systematic review and meta-analysis of evidence from epidemiological studies. **Accident Analysis and Prevention**, v. 60, p. 254–267, 2013.
- FERREIRA-VANEGAS, C. M.; VELEZ, J. I.; GARCIA-LLINAS, G. A. Analytical Methods and Determinants of Frequency and Severity of Road Accidents: A 20-Year Systematic Literature Review. **Journal of Advanced Transportation**, v. 2022, 2022.
- GARBARINO, S.; TRAVERSA, F.; SPIGNO, F. Sleepiness, safety on the road and management of risk. **Giornale Italiano di Medicina del Lavoro ed Ergonomia**, v. 34, n. 3, p. 322–325, 2012.
- GAURAV; ALAM, Z. Improving Road Safety in India Using Data Mining Techniques. In: PANDA, B.; SHARMA, S.; ROY, N. R. (Ed.). **Data Science and Analytics**. Anais... Singapore: Springer Singapore, 2018.
- GREGORIADES, A.; MOUSKOS, K. C. Black spots identification through a Bayesian Networks quantification of accident risk index. **Transportation Research Part C: Emerging Technologies**, v. 28, p. 28–43, 2013.
- HALAŠKA, M.; ŠPERKA, R. Performance of an automated process model discovery - The logistics process of a manufacturing company. **Engineering Management in Production and Services**, v. 11, n. 2, p. 106–118, 2019.
- HEVNER, Alan R.; MARCH, Salvatore T.; PARK, Jinsoo; RAM, Sudha. Design science in information systems research. **MIS Quarterly**, v. 28, n. 1, p. 75-105, 2004.
- KARYEMSETTY, N.; KUMAR, K. R. Road safety: An accident prevention using intelligent vehicular network. **International Journal of Safety and Security Engineering**, v. 10, n. 5, p. 631–638, 1 out. 2020.
- KASDE, R.; GUGAPRIYA, G. Accident avoidance system using CAN. **Indian Journal of Science and Technology**, v. 9, n. 30, 2016.

KEDEM-YEMINI, S.; MAMON, N. S.; MASHIAH, G. An Analysis of Cargo Release Services with Process Mining: A Case Study in a Logistics Company. Disponível em: <https://www.celonis.com>. Acesso em: 3 ago. 2022.

KHALIL, O. K. A Study on Road Accidents in Abu Dhabi Implementing a Vehicle Telematics System to Reduce Cost, Risk and Improve Safety. **Proceedings - International Conference on Developments in eSystems Engineering, DeSE**. Anais... 2018.

KIM, D. Development of practical guidelines for healthcare process mining. **ICIC Express Letters, Part B: Applications**, v. 9, n. 9, p. 977–982, 1 set. 2018.

KIZITO, A.; SEMWANGA, A. R. Modeling the Complexity of Road Accidents Prevention. **International Journal of System Dynamics Applications**, v. 9, n. 2, p. 24–41, 31 jan. 2020.

KURGANOV, V. et al. Process Mining as a Means of Improving the Reliability of Road Freight Transportations. **Transportation Research Procedia**. Anais... Elsevier B.V., 2021.

LACERDA, D. P.; DRESCH, A.; PROENÇA, A.; ANTUNES JÚNIOR, J. A. V. Design Science Research: método de pesquisa para a engenharia de produção. **Gestão & Produção**, v. 20, n. 4, p. 741-761, 2013. Disponível em: <https://www.scielo.br/j/gp/a/bhNdKfVCMT5vKTjqMBHJGgL/?lang=pt>. Acesso em: 20 out. 2023.

MAY, J. F.; BALDWIN, C. L. Driver fatigue: The importance of identifying causal factors of fatigue in traffic accidents. **Accident Analysis & Prevention**, v. 41, n. 5, p. 755-764, 2009.

MATEO, S. S. et al. Study of a driver assistance interface for merging situations on highways. **2018 IEEE International Conference on Vehicular Electronics and Safety, ICVES 2018**. Anais... Institute of Electrical and Electronics Engineers Inc., 31 out. 2018.

MESTRI, R. A.; RATHOD, R. R.; GARG, R. D. Identification and Removal of Accident-Prone Locations Using Spatial Data Mining. In: GHOSH, J. K.; SILVA, I. da (Ed.). **Applications of Geomatics in Civil Engineering**. Anais... Singapore: Springer Singapore, 2020.

MOHER, D. et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. **International Journal of Surgery**, v. 8, n. 5, p. 336–341, 2010.

NANDA, S.; SINGH, S. Evaluation of Factors Responsible for Road Accidents in India by Fuzzy AHP. In: PEREZ, G. M. et al. (Ed.). **Networking Communication and Data Knowledge Engineering**. Anais... Singapore: Springer Singapore, 2018.

OLARIU, C.; ORTEGA, J. D.; JAVIER YEBES, J. The role of cloud-computing in the development and application of ADAS. **European Signal Processing Conference**. Anais... European Signal Processing Conference, EUSIPCO, 29 nov. 2018.

OUAZZANI-TOUHAMI, K.; SOUISSI, N. Simulation project life cycle applied to road safety: A case study of morocco. **Colloquium in Information Science and Technology, CIST**. Anais... 2020.

PARKKARI, K.; PARKKARI, I.; SIHVOLA, N. 83 In-depth on-the-spot road accident investigation in Finland. **Injury Prevention**, v. 22, n. Suppl 2, p. A32, 1 set. 2016.

Plataforma Celonis. Disponível em: <https://www.celonis.com/>. Acesso em: 3 ago. 2022.

PEFFERS, Ken; TUUNANEN, Tuure; ROTHENBERGER, Marcus A.; CHATTERJEE, Samir. A design science research methodology for information systems research. **Journal of Management Information Systems**, v. 24, n. 3, p. 45-77, 2007.

RIAZ, F. et al. Internet of Things Based Driver Distraction Detection and Assistance System: A Novel Approach. **Ad Hoc & Sensor Wireless Networks**, v. 44, p. 305–336, 15 set. 2019.

RUDNITCKAIA, J.; INTAYOAD, W.; BECKER, T.; HRUSKA, T. Applying Process Mining to the Ship Handling Process at Oil Terminal. In: **INTERNATIONAL CONFERENCE ON INDUSTRY 4.0 AND PHYSICAL INTERNET (ICPS)**, 2019. p. 552-557. DOI: 10.1109/ICPHYS.2019.8780305.

SHAHZAD, M. Review of road accident analysis using GIS technique. **International Journal of Injury Control and Safety Promotion**, v. 27, n. 4, p. 472–481, 1 out. 2020.

Site Fluxicon - Disco. Disponível em: <https://fluxicon.com/disco/>. Acesso em: 3 ago. 2022.

STEVENSON, M. R. et al. The role of sleepiness, sleep disorders, and the work environment on heavy-vehicle crashes in 2 Australian states. **American Journal of Epidemiology**, v. 179, n. 5, p. 594–601, 1 mar. 2014.

TÖRÖK, Á.; PAUER, G. Assessment of the current status of Intelligent Transport Systems serving the improvement of road safety in Hungary. **Periodica Polytechnica Transportation Engineering**, v. 45, n. 2, p. 77–83, 2017.

UMA, S.; ESWARI, R. Accident prevention and safety assistance using IOT and machine learning. **Journal of Reliable Intelligent Environments**, v. 8, n. 2, p. 79–103, 1 jun. 2022.

VAN DER AALST, W. M. P.; WEIJTERS, A. J. M. M. Process mining: A research agenda. **Computers in Industry**, 2004.

VAN DER HEIJDEN, T. H. C. **Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice**. 2012. Tese (Mestrado em Engenharia Industrial) — Eindhoven University of Technology, Eindhoven, 2012.

VANLAAR, W.; YANNIS, G. Perception of road accident causes. **Accident Analysis and Prevention**, v. 38, n. 1, p. 155–161, jan. 2006.

VERGIDIS, K.; TURNER, C. J.; TIWARI, A. Business process perspectives: Theoretical developments vs. real-world practice. **International Journal of Production Economics**, v. 114, n. 1, p. 91–104, jul. 2008.

WANG, Y.; HULSTIJN, J.; TAN, Y. H. Regulatory supervision with computational audit in international supply chains. **ACM International Conference Proceeding Series**. Anais... Association for Computing Machinery, 30 maio 2018.

WORLD HEALTH ORGANIZATION. Global status report on road safety 2018. Geneva: World Health Organization, 2018. Disponível em: <https://www.who.int/publications-detail/global-status-report-on-road-safety-2018>. Acesso em: 22 ago 2022.

YUSOFF, N. M. et al. Selection of Measurement Method for Detection of Driver Visual Cognitive Distraction: A Review. **IEEE Access**, v. 5, p. 22844–22854, 2017.

ZUBAR, A.; SHABEER, A.; ZUBAR, H. A. Safety device to prevent cell phone accidents. **Journal of Scientific & Industrial Research**. Disponível em: <https://www.researchgate.net/publication/287040844>.

WIERINGA, Roel J. Design Science Methodology for Information Systems and Software Engineering. Springer, 2014. ISBN 978-3-662-43838-1.

WIERWILLE, W. W. Driver drowsiness detection: Analysis and modeling of driving behavior. National Highway Traffic Safety Administration, 1999.