



UNIVERSIDADE FEDERAL DO PARÁ  
CAMPUS ANANINDEUA  
FACULDADE DE TECNOLOGIA EM GEOPROCESSAMENTO

ADRIANA MEDEIROS PINHEIRO

**DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS DA POLÍCIA  
RODOVIÁRIA FEDERAL NOS ACIDENTES DA BR-316 NO ESTADO DO PARÁ  
UTILIZANDO A FERRAMENTA WEKA**

Ananindeua - PA  
2019

ADRIANA MEDEIROS PINHEIRO

**DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS DA POLÍCIA  
RODOVIÁRIA FEDERAL NOS ACIDENTES DA BR-316 NO ESTADO DO PARÁ  
UTILIZANDO A FERRAMENTA WEKA**

Trabalho de Conclusão de Curso apresentado à Faculdade de Tecnologia em Geoprocessamento da Universidade Federal do Pará, como requisito para obtenção do título de Tecnóloga em Geoprocessamento.

Orientadora: Dr<sup>a</sup>. Danielle Costa Carrara Couto  
Coorientador: Dr. Estêvão Jose da Silva Barbosa

ADRIANA MEDEIROS PINHEIRO

**DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS DA POLÍCIA  
RODOVIÁRIA FEDERAL NOS ACIDENTES DA BR-316 NO ESTADO DO PARÁ  
UTILIZANDO A FERRAMENTA WEKA**

Trabalho de Conclusão de Curso apresentado à Faculdade de Tecnologia em Geoprocessamento, do Campus de Ananindeua, da Universidade Federal do Pará, como requisito para obtenção do grau de Tecnólogo em Geoprocessamento.

Data da Aprovação: 10/01/2019

Conceito: Excelente

**BANCA EXAMINADORA:**

---

Prof<sup>a</sup>. Dr<sup>a</sup>. DANIELLE COSTA CARRARA COUTO  
Orientadora - UFPA

---

Prof. Dr. ESTÊVÃO JOSÉ DA SILVA BARBOSA  
Coorientador - UFPA

---

Prof. Ms. ALAN NUNES ARAÚJO  
Examinador Interno - UFPA

---

Prof<sup>a</sup>. Ms. CÁSSIA MARIA CARNEIRO KAHWAGE  
Examinadora Externa - Faculdade de Computação/UFPA

Dedico a todos meus futuros colegas da  
instituição Polícia Rodoviária Federal.

## AGRADECIMENTOS

A origem deste trabalho remonta ao ano de 2016, quando ingressei no curso de Tecnologia em Geoprocessamento. Desde então, fui diuturnamente perturbada por preocupações relativas a como poderia ajudar a segurança pública do meu estado na prevenção de acidentes de trânsito usando técnicas de geoprocessamento e o auxílio da mineração de dados na otimização da análise dos acidentes rodoviários. Todo o estudo desenvolvido procura auxiliar na tomada de decisão. Felizmente, não estive sozinha em minha jornada. Contei com o auxílio de futuros colegas de valor de carreira militar e policial como o Adriano Ferreira da Polícia Rodoviária Federal, Major Pires que trabalha em Brasília coordenando projetos de mapeamento do exército brasileiro, bem como a Major Soraia. O apoio concedido por eles e pelo meu melhor amigo Sasayamomo que mora em Taiwan no início da construção deste trabalho, foi muito importante para a conclusão do mesmo.

Cabe referência especial a professora Danielle Couto por sua orientação e ao professor Estêvão Barbosa por suas instruções como co orientador. Sou grata, também, a todos os professores da Faculdade de Tecnologia em Geoprocessamento que me instruíram ao longo do curso. Serei eternamente grata por cada gesto. Também tive o privilégio de ter sido ouvida por alguns professores como Rodrigo Quites e Marcos Seruffo, ambos da UFPA - Campus Universitário de Belém, os quais foram extremamente prestativos com a minha pessoa. Não poderia, ainda, deixar de agradecer aos meus pais pela boa educação. Menção especial merece o meu irmão gêmeo Adriano por sempre estar sonhando os meus sonhos e, por fim e, mais importante, sou grata ao meu Deus que me concedeu sabedoria e cuidou de cada detalhe do desenvolvimento deste trabalho. Não há aqui espaço suficiente para expor meus sentimentos por Ele.

"Se eu conseguisse enxergar mais longe, é porque procurei estar sobre ombros de gigantes."  
(ISAAC NEWTON, 1676).

## RESUMO

Um dos grandes desafios da era da informação é transformar informações em conhecimento. Neste contexto, houve um progresso impressionante em mineração de dados e aprendizado de máquina nos últimos anos. A aprendizagem de máquina, síntese de estatísticas, teoria da informação e computação criou uma ciência sólida, com uma base matemática firme e com ferramentas muito poderosas buscando solucionar este desafio. Diante disso, a mineração de dados surge como extração de informações implícitas, anteriormente desconhecidas e potencialmente úteis. O presente trabalho teve como objetivo principal usar o banco de dados do Departamento de Trânsito da Polícia Rodoviária Federal e construir associações de acidentes de trânsito buscando regularidades ou padrões, usando o software livre Weka. Foi usado dados de acidentes da BR-316 no estado do Pará entre Janeiro de 2007 a Setembro de 2018. A descoberta de conhecimento se deu por meio de regras de associações geradas pelo software livre Weka através do algoritmo *Apriori*. Por meio de resultados estatísticos do programa gerou-se gráficos que mostram fatores que mais contribuíram para os acidentes de trânsito na rodovia BR-316 na última década. Os resultados apontam que nos quilômetros iniciais da rodovia capitão Pedro Teixeira (BR-316) no Pará e mais precisamente nos quilômetros 0 e 8, em pleno dia, e nos municípios Ananindeua, Marituba e Castanhal há um alto índice de acidentes de trânsito.

**Palavras-chave:** Acidentes de Trânsito. Mineração de Dados. Weka.

## ABSTRACT

One of the great challenges of the information age is to transform information into knowledge. In this context, there has been impressive progress in data mining and machine learning in recent years. Machine learning, synthesis of statistics, information theory and computation has created a solid science with a firm mathematical base and with very powerful tools seeking to solve this challenge. Given this, data mining appears as the extraction of implicit information, previously unknown and potentially useful. The main objective of this work was to use the database of the Traffic Department of the Federal Highway Police and to build associations of traffic accidents seeking regularities or standards, using the free software Weka. We used accident data from BR-316 in the state of Pará between January 2007 and September 2018. The knowledge discovery occurred through association rules generated by the free software Weka through the Apriori algorithm. The program's statistical results generated graphs that show the factors that contributed the most to traffic accidents on the BR-316 highway in the last decade. The results indicate that in the initial kilometers of the captain Pedro Teixeira highway (BR-316) in Pará and more precisely in kilometers 0 and 8, in the daytime, and in the municipalities of Ananindeua, Marituba and Castanhal there is a high rate of traffic accidents.

**Keywords:** Transit Accidents. Data Mining. Weka.

## LISTA DE FIGURAS

<b>Figura 1</b> - Objetivo da Década de Ação pela Segurança no Trânsito.....	13
<b>Figura 2</b> – Etapas do processo do KDD.....	14
<b>Figura 3</b> - As dez principais causas de morte entre jovens de 15 a 29 anos em 2012.....	17
<b>Figura 4</b> - Sistema Seguro de Segurança Rodoviária.....	20
<b>Figura 5</b> - Evolução da Geoinformação.....	21
<b>Figura 6</b> - Etapas de Produção da Geoinformação.....	22
<b>Figura 7</b> - Encontro das Rodovias BR-316 e BR-010 no estado do Pará.....	24
<b>Figura 8</b> - Rodovias Federais do Pará.....	25
<b>Figura 9</b> - Rodovia BR-316 no Estado do Pará.....	32
<b>Figura 10</b> - Página do Departamento da Polícia Rodoviária Federal (Datatran/DPRF).....	33
<b>Figura 11</b> - Tela Inicial do Weka.....	34
<b>Figura 12</b> - Exemplo da extração.....	36
<b>Figura 13</b> - Janela de pré-processamento no Weka.....	37
<b>Figura 14</b> - Aplicação do filtro NumericToNominal.....	37
<b>Figura 15</b> - Conversão do Atributo Horário.....	38
<b>Figura 16</b> - Distribuição do Horário após a aplicação do Filtro mergeManyValues.....	38
<b>Figura 17</b> - Distribuição do atributo Km antes de ser usado o filtro Discretize.....	39
<b>Figura 18</b> - Estrutura final do arquivo ARFF.....	40
<b>Figura 19</b> - Distribuição anual dos acidentes da BR-316 no estado do Pará.....	41
<b>Figura 20</b> - Distribuição mensal dos Acidentes de Trânsito no estado do PA entre os anos de 2007 a setembro de 2018.....	42
<b>Figura 21</b> - Quantidade de ferido graves dos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	43
<b>Figura 22</b> - Classificação dos Trechos das Rodovias do Pará.....	43
<b>Figura 23</b> - Distribuição total de Mortes das Rodovias Federais do Pará entre Janeiro de 2007 a Setembro de 2018.....	44
<b>Figura 24</b> - Traçado da Via nos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	45
<b>Figura 25</b> - Condição Meteorológica nos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	45
<b>Figura 26</b> - Classificação dos Acidentes de trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	46
<b>Figura 27</b> - Tipo de Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	46
<b>Figura 28</b> - Causa dos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	47
<b>Figura 29</b> - Municípios dos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	48
<b>Figura 30</b> - Horário dos acidentes de trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	49
<b>Figura 31</b> - Dia da semana dos acidentes de trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	49
<b>Figura 32</b> - Países com legislação de velocidade que cumprem as melhores práticas, 2017.....	50
<b>Figura 33</b> - Fotografia no km 8 da Rodovia BR-316.....	50
<b>Figura 34</b> - Distribuição dos Quilômetros antes da aplicação do filtro Discretize.....	51
<b>Figura 35</b> - Distribuição dos quilômetros dos acidentes de trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	51
<b>Figura 36</b> - Teste de algoritmo Apriori no Weka com dados de Acidente de Trânsito da Polícia Rodoviária Federal na BR-316.....	52
<b>Figura 37</b> - Parâmetro padrão do algoritmo Apriori no Weka.....	53
<b>Figura 38</b> - Cenário 1 – atributos: horário, km, município e causa do acidente.....	54
<b>Figura 39</b> - Cenário 2 – atributos: dia da semana, município, causa do acidente e tipo de acidente.....	56

<b>Figura 40</b> - Dia do mês dos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.....	66
<b>Figura 41</b> - Sentido da Via nos Acidentes de Trânsito da BR-316 no estado do PA.....	66
<b>Figura 42</b> - Fase do dia nos Acidentes de Trânsito da BR-316 no estado do PA.....	66

## LISTA DE TABELAS

<b>Tabela 1</b> – Rodovias Federais do Pará e suas extensões.....	24
<b>Tabela 2</b> – Tarefas e Técnicas de Mineração de Dados.....	26
<b>Tabela 3</b> - Tipos de Regras de associação no software Weka.....	28
<b>Tabela 4</b> – Metodologia utilizada na Pesquisa.....	31
<b>Tabela 5</b> – Quantidade de Registros de Acidentes de Trânsito Processados.....	35
<b>Tabela 6</b> - Dicionário de Dados.....	36
<b>Tabela 7</b> - Quantidade total de Mortes das Rodovias do Pará entre Janeiro de 2007 a Setembro de 2018.....	44
<b>Tabela 8</b> – Regras de Associação encontradas no cenário 1.....	55
<b>Tabela 9</b> - Regras de Associação encontradas no cenário 2.....	56

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>13</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>17</b>
<b>2.1</b>	<b>Acidentes de trânsito e segurança rodoviária.....</b>	<b>17</b>
<b>2.2</b>	<b>Geoinformação como instrumento de análise aos acidentes rodoviários.....</b>	<b>21</b>
<b>2.3</b>	<b>A Rodovia BR-316.....</b>	<b>22</b>
<b>2.4</b>	<b>Mineração de dados na análise de acidentes rodoviários.....</b>	<b>25</b>
2.4.1	Regras de Associação.....	27
<b>2.5</b>	<b>Trabalhos correlatos.....</b>	<b>29</b>
<b>3</b>	<b>MATERIAIS E MÉTODOS.....</b>	<b>31</b>
<b>3.1</b>	<b>Banco de dados da polícia rodoviária federal.....</b>	<b>32</b>
<b>3.2</b>	<b>Software weka.....</b>	<b>33</b>
<b>3.3</b>	<b>Procedimentos metodológicos.....</b>	<b>34</b>
3.3.1	Obtenção de Dados.....	34
3.3.2	Pré-processamento dos Dados.....	35
3.3.3	Transformação dos Dados.....	39
<b>4</b>	<b>RESULTADOS E DISCUSSÕES.....</b>	<b>41</b>
<b>4.1</b>	<b>Análise dos dados estatísticos gerais dos acidentes de trânsito da BR-316 no estado do Pará entre Janeiro de 2007 a Setembro de 2018.....</b>	<b>41</b>
<b>4.2</b>	<b>Resultados com o algoritmo Apriori envolvendo os atributos horário, km, município e causa do acidente.....</b>	<b>52</b>
<b>4.3</b>	<b>Resultados com o algoritmo Apriori envolvendo os atributos dia da semana, município, causa do acidente e tipo de acidente.....</b>	<b>55</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>58</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>59</b>
	<b>APÊNDICES.....</b>	<b>63</b>
	<b>APÊNDICE A - INFORMAÇÕES DAS PLANILHAS BAIXADAS NO BANCO DE DADOS DA PRF.....</b>	<b>64</b>
	<b>APÊNDICE B – ESTRUTURA USADA NO ARQUIVO ARFF.....</b>	<b>65</b>
	<b>APÊNDICE C – OUTROS GRÁFICOS GERADOS NO WEKA.....</b>	<b>66</b>

## 1 INTRODUÇÃO

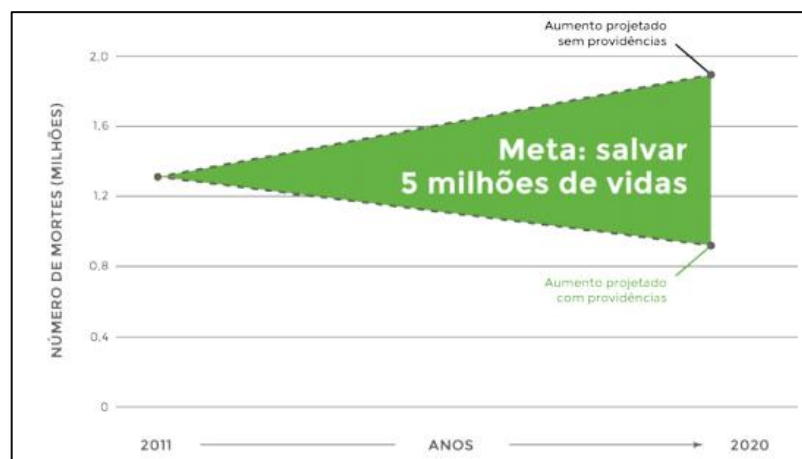
A inteligência artificial (IA) está alterando o mundo e levantando questões importantes para a sociedade, a economia e a governança. Ela geralmente é realizada em conjunto com aprendizagem de máquina e análise de dados. A aprendizagem de máquina coleta dados e procura tendências subjacentes. Se ela encontrar algo relevante para um problema prático, os projetistas de software podem usar esse conhecimento e usá-lo para analisar problemas específicos. Para uma boa descoberta de conhecimento são necessários dados suficientemente robustos para que os algoritmos possam discernir padrões úteis. Os dados podem vir na forma de informações digitais, imagens de satélite, informações visuais, texto ou dados não estruturados (WEST;ALLEN,2018).

Diante disso, os acidentes de trânsito se configuram como grave problema de saúde pública de um País. As estatísticas mostram que o trânsito brasileiro é o quarto mais violento do continente americano, segundo dados divulgados pela Organização Mundial da Saúde (OMS, 2015).

No ano de 2013, cerca de 41 mil pessoas foram mortas por causa de acidentes nas vias brasileiras. No período de 2009 a 2013 o número de acidentes de trânsito no Brasil teve um aumento de 19% para 23,4% por 100 mil habitantes. Este foi o maior registro da América do Sul, segundo informações de um relatório divulgado pela Organização das Nações Unidas no Brasil (ONUBR, 2011).

Este fenômeno é hoje mundialmente preocupante. As Nações Unidas, através de Resolução A/RES/64/255, publicada no dia 02 de Março de 2010, proclamou o período de 2011 a 2020 como a “Década de Ações para a Segurança no Trânsito” (Figura 1).

**Figura 1** - Objetivo da Década de Ação pela Segurança no Trânsito

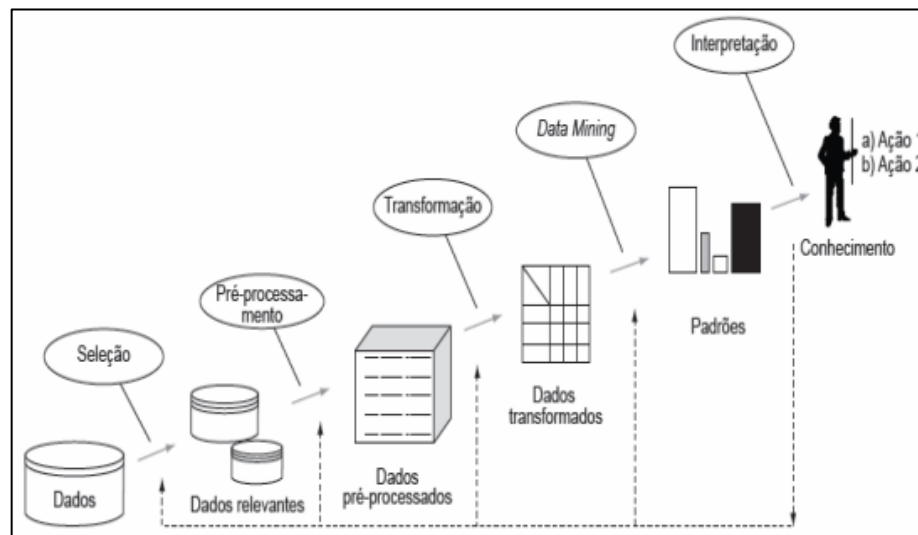


**Fonte:** Organização Mundial da Saúde, 2013.

Nessa perspectiva, a mineração de dados pode ajudar na tomada de decisões descobrindo conexões escondidas e prevendo tendências futuras da indústria privada, administração pública, forças armadas e/ou saúde. Neste trabalho, tomou-se como área de estudo um trecho de 170 km da Rodovia BR-316 que percorre a região Norte do Brasil, no estado do Pará, entre os municípios de Belém (capital) e Viseu (divisa com o Maranhão).

Para auxiliar na identificação de padrões dos acidentes rodoviários, será feito uso da descoberta de conhecimento em banco de dados (*Knowledge Discovery in Database – KDD*) (Figura 2), com a técnica de mineração de dados e o algoritmo *Apriori* pela ferramenta livre *Waikato Environment for Knowledge Analysis (WEKA)*.

**Figura 2** – Etapas do processo do KDD



**Fonte:** GALVÃO. N. D, 2009.

A mineração de dados rodoviários na prevenção de acidentes de trânsito é importante por ser responsável pela automação da maioria dos métodos de aprendizado de máquina na realização de tarefas especializadas, pois este processo descobre informações relevantes, como padrões, associações, mudanças, anomalias e estruturas em grande quantidade de dados armazenados em banco de dados.

Bastos (2011) assegura que, para uma correta avaliação da segurança viária no Brasil, a mensuração ou quantificação da acidentalidade no trânsito é muito importante. Com valores numéricos, é possível comparar a situação da acidentalidade em diferentes locais, como países, estados, municípios e rodovias, ou mesmo em espaços geográficos, em diferentes épocas, o que é imprescindível na avaliação de ações voltadas para a redução de acidentes e de mortes e feridos no trânsito.

Partindo do que foi observado em estudos nacionais e internacionais sobre o tema, e levando em consideração a utilização de técnicas de mineração de dados na construção de diagnósticos, a presente pesquisa tem como intuito analisar as condições que mais contribuíram para registros de acidentes na última década e identificar padrões de acidentes de trânsito na rodovia BR-316 no estado do Pará usando a ferramenta WEKA - plataforma livre de aprendizado de máquina que fornece uma interface gráfica interativa, um conjunto de algoritmos de última geração e muito propícia para iniciante em aprendizado de máquina aplicado.

A pesquisa tem como objetivo principal efetuar o diagnóstico dos acidentes rodoviários ocorridos na BR-316 no estado do Pará, através de técnicas de mineração de dados das ocorrências de acidente de trânsito verificadas entre Janeiro de 2007 a Setembro de 2018.

Os objetivos específicos:

- Levantar e organizar dados de acidentes de trânsito ocorridos na BR-316 no período de janeiro de 2007 a setembro de 2018 no estado do Pará, provenientes da base de dados online da Polícia Rodoviária Federal;
- Verificar os padrões de associação de variáveis nos acidentes rodoviários na BR-316 utilizando o software WEKA;
- Analisar fatores que contribuíram com os acidentes observados na BR-316 do estado do Pará a partir de algoritmos de Mineração de Dados;
- Apresentar por meio de Tabelas, gráficos e mapas temáticos os resultados obtidos;
- Tirar ilações para a tomada de decisões para que órgãos de gestão do sistema rodoviário possam utilizar os resultados obtidos.

Julga-se pertinente a análise, uma vez que, existem poucos trabalhos no estado do Pará que procuram ajudar na gestão das políticas públicas na tomada de decisões para mitigar os acidentes de trânsito na rodovia BR-316. A análise efetuada e o conhecimento alcançado a respeito deste assunto, poderá servir de alerta ou guia aos órgãos gestores do sistema viário na planificação de ações adequadas objetivando reduzir significativamente esses índices.

Este trabalho está organizado em seis seções. O primeiro capítulo consiste na introdução do trabalho. A seção 2 apresenta o referencial teórico, explicando as áreas de conhecimento utilizadas para a compor o presente trabalho; a seção 3 descreve os três (3) trabalhos correlatos que norteram nossa pesquisa; na seção 4 os materiais e métodos, demonstrando os procedimentos metodológicos aplicados no trabalho; a seção 5 apresenta os resultados obtidos a partir da mineração de dados realizados com o software WEKA utilizando

o banco de dados da Polícia Rodoviária Federal; e por fim, a seção 6 contém as considerações finais, bem como informando os trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Acidentes de trânsito e segurança rodoviária

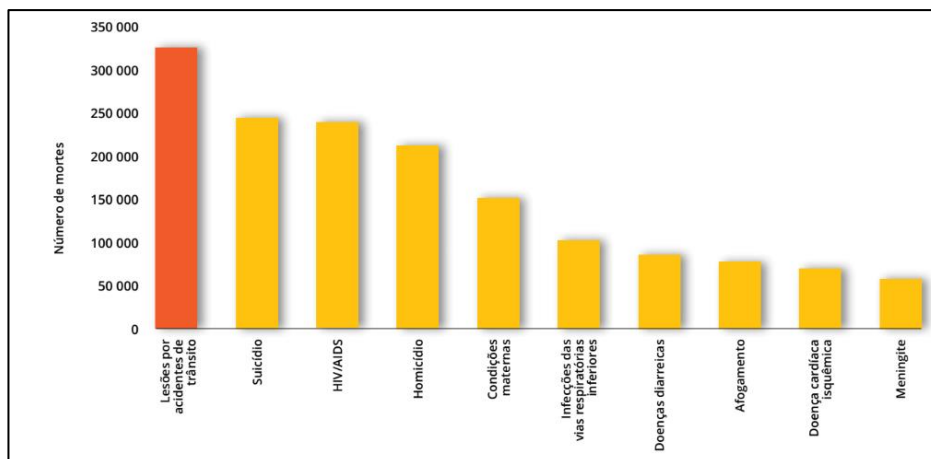
Após o fim da Segunda Guerra Mundial, teve início um processo de aumento da utilização dos automóveis nos grandes centros urbanos que modificou o desenho das cidades e o padrão de deslocamento nesses espaços. Uma das diversas consequências negativas da consolidação do automóvel como principal meio de transporte nas metrópoles foi o aumento do número de acidentes no trânsito (SILVA, 2018).

No início, o automóvel era um objeto de luxo, mas, de repente, tornou-se um utensílio essencial da modernidade. Essa rápida mudança teve consequências sobre a sociedade, como a mortalidade causada por acidentes de trânsito. As taxas de mortalidade atingiram números alarmantes para a sociedade moderna, e neste contexto muitos estudos, projetos, iniciativas e campanhas têm sido realizados a fim de minimizar o problema e maximizar as soluções.

Os acidentes de trânsito constituem um sério e grave problema no mundo contemporâneo. Essa situação levou a ONU e as nações se comprometerem em adotar novas medidas para prevenir e reduzir o número de mortes em acidentes viários (ONUBR, 2011).

De acordo com o relatório global de 2015 sobre o estado da segurança viária, divulgado pela OMS, lesões por acidentes de trânsito ocuparam o primeiro lugar das 10 principais causas de mortes entre jovens de 15 a 29 anos em 2012 (Figura 3).

**Figura 3** - As dez principais causas de morte entre jovens de 15 a 29 anos em 2012



**Fonte:** OMS, 2015.

Explicar como se deu um acidente viário é um grande desafio desde que a primeira ocorrência foi registrada, em Londres, no ano de 1909. Nas primeiras abordagens do assunto, o acidente era tratado como uma ocorrência aleatória sobre os quais o ser humano não teria controle algum. Em 1920, durante a Primeira Guerra Mundial, surgiu a teoria da propensão com

a ideia de que só ocorreriam acidentes desde que existissem certas tendências iniciais. Na década de 1950, estudos evidenciaram que a prevenção dos acidentes seria possível, que haveria sucesso nessa prevenção, e que ela só seria alcançada se as causas do provável acidente fossem determinadas (BASTOS, 2011).

Entre 1960 e 1970, popularizou-se a teoria epidemiológica<sup>1</sup>, considerando que os acidentes seriam resultados de desajustes na interação entre várias partes dos sistemas complexos, modificando os componentes técnicos do sistema de transporte, principalmente, aqueles relacionados à engenharia de segurança (ELVIK et al., 2009).

De fato, para a comunidade científica, ainda não há teoria ideal. A abordagem mais tradicional tem base a estatística de frequência de ocorrência e na severidade dos acidentes (HAUER, 2007). Isso porque os especialistas descobriram que os acidentes não são acontecimentos fortuitos, imprevisíveis e dependentes da sorte, mas, ao contrário, eles seguem parâmetros característicos de distribuição, isto é, o acidente é sempre uma consequência (HOFFMANN, 2005). A segurança de uma rede rodoviária está relacionada principalmente com a quantidade de acidentes graves.

No Brasil, a primeira norma legal sobre o trânsito, de que se tem conhecimento, é o Decreto nº8324 de 27 de outubro de 1910, cuja principal finalidade era disciplinar o serviço de transporte por automóvel. Atualmente, o Código de Trânsito Brasileiro (CTB), em vigor desde 23 de setembro de 1998, impõe padrões de controle bem mais severos que os anteriores (MOUKARZEL, 1999).

No relatório da OMS, do ano de 2015, o Brasil ocupava o 5º lugar no ranking do número de mortes no trânsito no mundo, posicionando-se atrás apenas da Índia, China, EUA e Rússia. Sem campanhas de conscientização, a OMS estima que 1,9 milhão de pessoas devem morrer no trânsito em 2020 (passando para a quinta maior causa de mortalidade) e 2,4 milhões em 2030.

No livro “*Volunteered Geographic Information and the Future of Geospatial Data*” (2017), cujo autores Cláudio Elízio Calazans Campelo (Universidade Federal de Campina Grande, Brazil), Michela Bertolotto (University College Dublin, Ireland) and Pdraig Corcoran (Cardiff University, UK) mostram que medidas preventivas são melhores que correções reativas, uma vez que nos métodos reativos o número de fatalidades e pessoas gravemente feridas é usado como um indicador para determinar se uma seção da via precisa ser

---

<sup>1</sup> Epidemiologia é o estudo da frequência, da distribuição e dos determinantes dos problemas de saúde em populações humanas, bem como a aplicação desses estudos no controle dos eventos relacionados com saúde.

investigada ou não, ou seja, é necessário que ocorra mortes no trânsito para fazer tomadas de decisões.

Métodos proativos, por outro lado, tentam determinar pontos perigosos e consertá-los antes que ocorra óbitos na seção. Um exemplo de medida preventiva, seria inspeções de segurança rodoviária (ISR) - monitoramento realizado regularmente das estradas existentes com o objetivo de identificar os riscos de tráfego relacionados às características do ambiente rodoviário e propor medidas para mitigar os perigos detectados, como foi o caso do projeto *Ripcord-Iserest* da União Europeia, com participação de 14 países, no qual foi realizado um inquérito para determinar as melhores práticas para o ISR com modelos de previsão de acidentes.

Uma boa visão geral do *Ripcord-Iserest* (ISR) praticado na Europa é dada em seu manual por Nadler, Nadler e Strnad (2014), os quais afirmam que um ISR pode ser iniciado por:

- Quando uma concentração de acidentes é observada, especialmente com estruturas de tipos diversos;
- Quando evidências, problemas ou outras informações tornam uma inspeção prudente;
- Quando existem deficiências de segurança, potencial de risco ou os mesmos tipos de acidentes em seções mais longas.

O projeto alcançou resultados satisfatórios como (NADLER; NADLER; STRNAD, 2014):

1. Um levantamento aprofundado da literatura de Modelos de Previsão de Acidentes (MPA) e de Avaliação de Impacto na Segurança Rodoviária. O projeto concluiu que o MPA pode diferir para o mesmo tipo de estrada em diferentes países, abrindo caminho para novas recomendações sobre o modo como esses instrumentos podem ser usados pelos praticantes.
2. Um foco no conceito de "estradas auto-explicativas", uma estrada projetada e construída de forma a induzir um comportamento adequado em nome do motorista. O projeto encaminhou recomendações de projeto de estradas relativas ao *layout* e elementos de projeto de estradas secundárias, incorporando modelos de comportamento do usuário da estrada.
3. O projeto desenvolveu o SEROES, um banco de dados de acesso livre para autoridades rodoviárias responsáveis por estradas secundárias, contendo o "Sistema de Informações de Segurança de Melhores Práticas", um resumo de informações sobre melhorias na

segurança viária. Em mais uma etapa, o SEROES foi demonstrado em conjunto com a Ferramenta de Segurança de Suporte à Decisão.

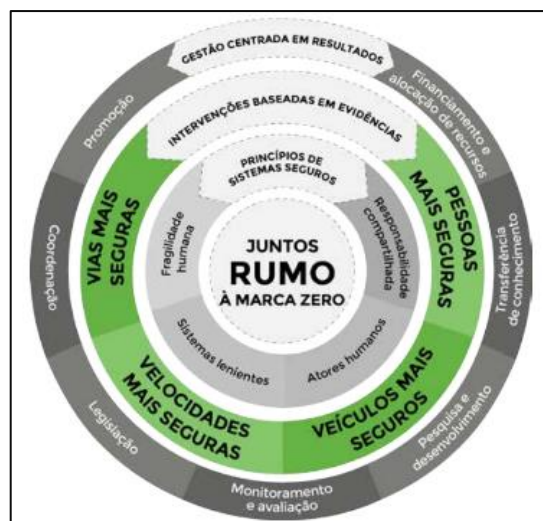
4. Uma Ferramenta de Apoio à Decisão em Sistemas de Informação Geográfica (SIG) foi desenvolvida para ajudar as autoridades rodoviárias locais e regionais a determinar problemas de segurança no trânsito, selecionar medidas apropriadas de intervenção de segurança e fornecer a oportunidade de prever o nível de segurança rodoviária de um município/região e a relação custo-eficácia dos projetos de segurança viária.

Em termos gerais, o projeto ISR, através das suas recomendações e desenvolvimentos técnicos, contribuiu para estabelecer normas e ferramentas práticas para os gestores de estradas locais europeus.

Além disso, o projeto estabeleceu recomendações em um Manual de Segurança para Estradas Secundárias, delineando os aspectos organizacionais e de implementação da segurança no trânsito. Do ponto de vista da política local, as ferramentas do projeto permitem que os resultados de segurança viária sejam comparados com as metas de segurança rodoviária feitas por formuladores de políticas regionais ou nacionais, permitindo assim um “aperfeiçoamento” do projeto de infraestrutura de segurança viária e comparações em toda a Europa (CORDIS, 2010).

Um país que contém um sistema seguro de segurança rodoviária (Figura 4) fica em uma melhor posição para vencer na competição global por investimentos econômicos e tem mais probabilidade de obter retornos elevados sobre os gastos com educação e saúde alinhados aos limites de velocidade e comportamentos do usuário do trânsito às qualidades de proteção de vias e veículos.

**Figura 4** - Sistema Seguro de Segurança Rodoviária



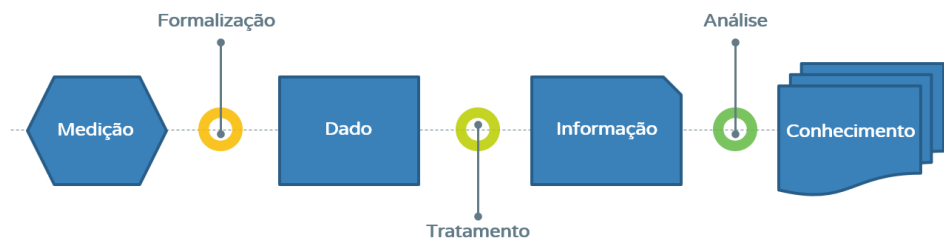
**Fonte:** Governo da Austrália Meridional (2011).

## 2.2 Geoinformação como instrumento de análise aos acidentes rodoviários

Acidentes rodoviários são complexos para analisar, uma vez que atravessa as fronteiras da engenharia, geografia e comportamento humano (Mahmud 2010). No Brasil, a ocorrência de acidentes rodoviários é registrada em boletins de ocorrência (BO) por agentes das polícias federal e estadual. Tais registros são quantitativos e qualitativos, em modo textual, porém são dados passíveis de espacialização.

A Geoinformação (Figura 5) é um conhecimento associado a uma referência geográfica – ou georreferenciamento – resultado do processamento de dados espaciais, que se destina a apoiar o processo decisório, e tem o objetivo de estudar e implementar diferentes formas de representação computacional do espaço geográfico (Manual de Geoinformação do Exército Brasileiro, 2014).

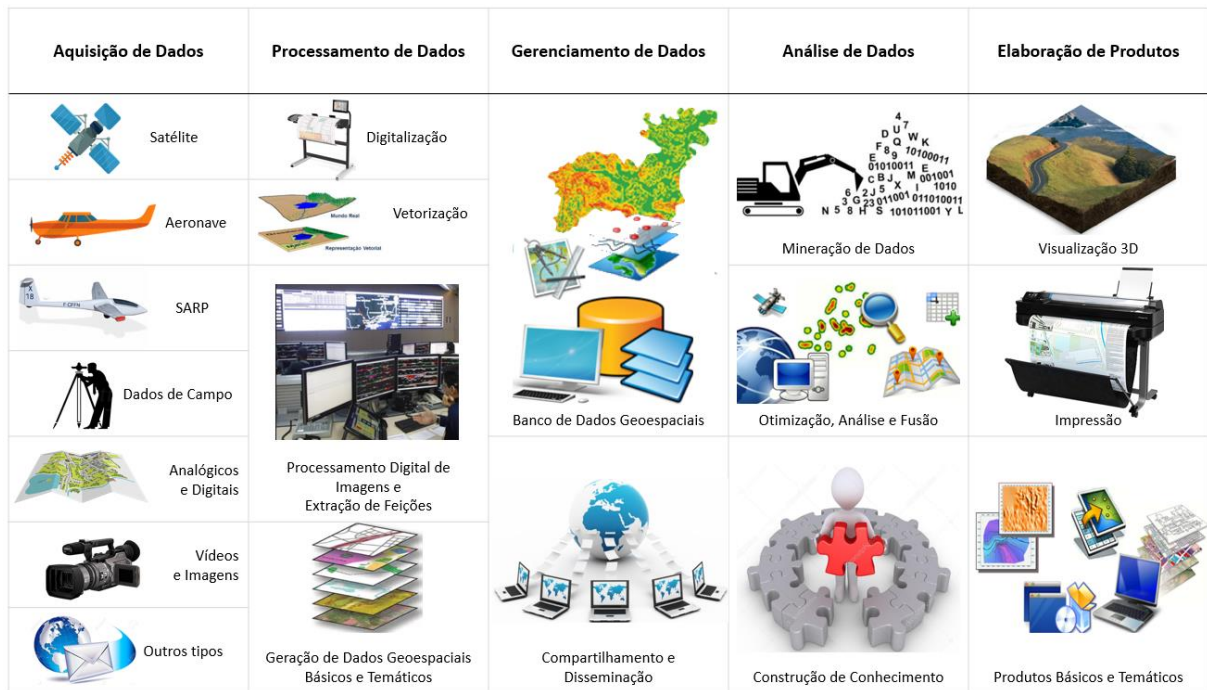
**Figura 5** - Evolução da Geoinformação



**Fonte:** Adaptado do Manual de Geoinformação do Exército Brasileiro, 2014.

Ela é composta por geotecnologias, processo de produção (metodologias), produtos (básicos e temáticos), e infraestrutura (hardware e software). Sendo assim, ela pode ser considerada um campo de análise dos acidentes rodoviários, haja vista que possui capacidades como integração para atividades e tarefas de comando, controle, inteligência; gestão de grandes quantidades de dados (*Big Data*); proteção cibernética; fusão de dados; computação em nuvem (*Cloud Computing*); e interoperabilidade.

As ferramentas da Geoinformação são recursos expressivos para se atuar adequadamente nos planejamentos espaciais. Se incorporados a uma base de dados geográficos podem ser representados em um mapa associado a outros formatos de representação e assim, visualizar os trechos das rodovias com maior ou menor incidência de ocorrências. As etapas de produção da Geoinformação podem ser divididas em cinco etapas como mostra a Figura 6.

**Figura 6 - Etapas de Produção da Geoinformação**

**Fonte:** Adaptado do Manual de Geoinformação Exército Brasileiro, 2014.

A Geoinformação vem sendo aplicada na combinação de tecnologia de geoinformação e análise espacial estatística para destacar a influência de fatores espaciais em sua formação. Shafabakhsh, Famili e Bahadori (2017) aplicaram essa combinação e fizeram 4 análises de agrupamento para uma melhor compreensão dos padrões de acidentes de trânsito em redes urbanas complexas em Mashhad, Nordeste do Irã, utilizando SIG como ArcMap e SANET.

### 2.3 A RODOVIA BR-316

Juscelino Kubitschek ao governar o país em 1956, comprometeu-se a trazer desenvolvimento, realizando 50 anos de progresso em 5 anos de governo. Seu Plano de Metas privilegiava o setor de bens de consumo sofisticado (automóvel, produtos eletrônicos, etc.), e bens intermediários (combustíveis líquidos, siderurgia, carvão, etc.), além de criar infraestrutura adequada para o engajamento maior do setor privado em setores mais avançados de industrialização, reforçando assim o papel da urbanização como base para a industrialização (ARENDA, 2009; BECKER e EGLER, 1998; OLIVEIRA, 2009).

Os investimentos estatais em infraestrutura tinham seis áreas gerais: energia, transporte, fornecimento de alimentação, indústria de base, educação e a construção de Brasília. Nos transportes, JK deixou de investir no segmento ferroviário para o rodoviário, pois objetivava introduzir o setor automobilístico no país (REGO; MARQUES, 2006; RODRIGUES, 2009; SCARLATO, 2009; TAFFAREL, 2010).

A ênfase dada ao setor de transporte foi consolidada através do Plano Nacional da Indústria Automobilística (1956), que atraiu multinacionais automobilísticas pelas suas vantagens. A partir da criação do Grupo Executivo da Indústria Automobilística (GEIA) implantaram-se várias industriais de automóveis no país (Ford, Volkswagen, General Motors, Fiat, etc.), influenciando o governo na criação de novas rodovias, como a rodovia Bernardo Sayão, mais conhecida como Belém-Brasília (BR-010) inaugurada em 1960, o primeiro elo rodoviário da região amazônica com o centro-sul com 1.959 quilômetros de extensão.

A rodovia homenageia o engenheiro Bernardo Sayão que morreu atingido pela queda de uma árvore em 15 de janeiro de 1959 durante a construção da rodovia (AREND, 2009; BECKER, 2009; FERREIRA & DIAS, 2001; FERREIRA, 2005; MEIRELLES FILHO, 2006; MENEZES, 2007; OLIVEIRA, 2009; REGO & MARQUES, 2006; RODRIGUES, 2009; SCARLATO, 2009).

Nesse contexto, configurou-se a malha rodoviária brasileira formada por rodovias federais e estaduais construídas por empreiteiras paulistas e mineiras, como a Camargo Corrêa, Andrade Gutierrez e a Mendes Junior, sendo estas as maiores parceiras das grandes obras do governo militar.

A identificação da malha federal utilizou-se do prefixo BR, dividida em: Radiais (começam em Brasília, numeradas de 1 a 100), Longitudinais (sentido Norte-Sul, numeradas de 101 a 200), Transversais (sentido Leste-Oeste, numeradas de 201 a 300), Diagonais (sentido diagonal, numeradas de 301 a 400) e Ligação (unem as anteriores, numeradas de 401 a 500) (MEIRELLES FILHO, 2006; RODRIGUES, 2009).

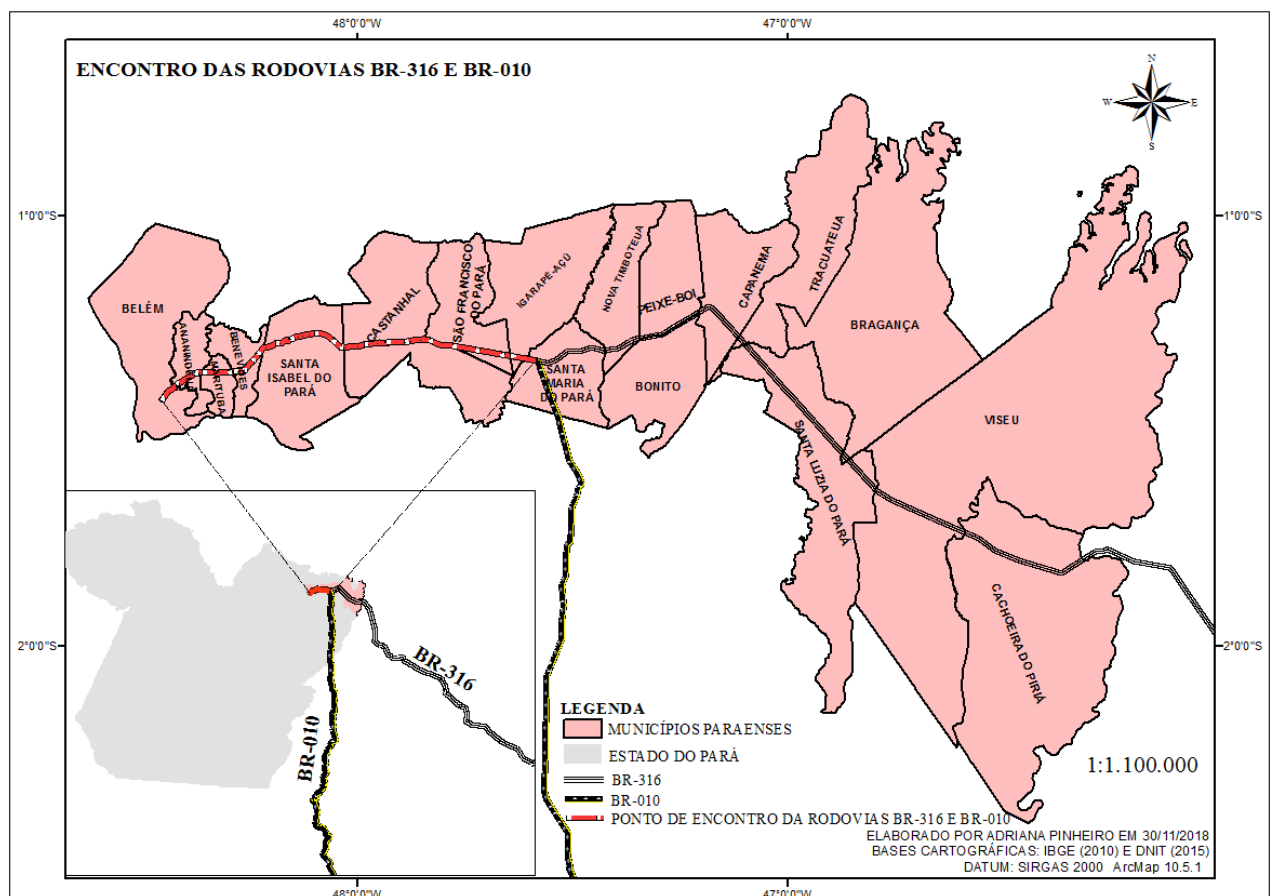
A rodovia capitão Pedro Teixeira (BR-316) tem seu início no estado do Pará, mais precisamente no município de Belém, e atravessa o Maranhão, Piauí, Pernambuco e Alagoas, e possui uma extensão de aproximadamente 2.067 km.

Com cinco décadas de existência, a Rodovia BR-316 é precária, tanto pela conservação, como pela obstrução em diversos pontos devido à ocupação não planejada de suas margens, principalmente a partir da década de 1970, quando muitas empresas comerciais se instalaram nos primeiros quilômetros (SETRAN, 2018).

A BR-010 e a BR-316 se unem em um único elo rodoviário no município de Santa Maria do Pará, percorrendo mais de 100 km até chegar à capital paraense, como mostra a Figura 7. A BR-316 tem seu Km 0 no Entroncamento no bairro do Castanheira, enquanto a BR-010 tem seu término em São Braz no fim da Avenida Almirante Barroso, próximo ao terminal rodoviário. Hoje, esse elo mantém uma função comercial com várias empresas de automóveis, além de redes de supermercados, lojas, bancos, farmácias, hospitais, lojas de materiais de construção e residenciais horizontais e verticais, configurando-se como um dos trechos mais urbanizados do Pará.

A Tabela 1 e Figura 8 mostram as rodovias federais do Pará e suas extensões.

**Figura 7** - Encontro das Rodovias BR-316 e BR-010 no estado do Pará



Fonte: Autora.

**Tabela 1** – Rodovias Federais do Pará e suas extensões

BR	Extensão Total (Km)
010	2.065,20
153	3.677,10
155	343,90
158	4.326,50

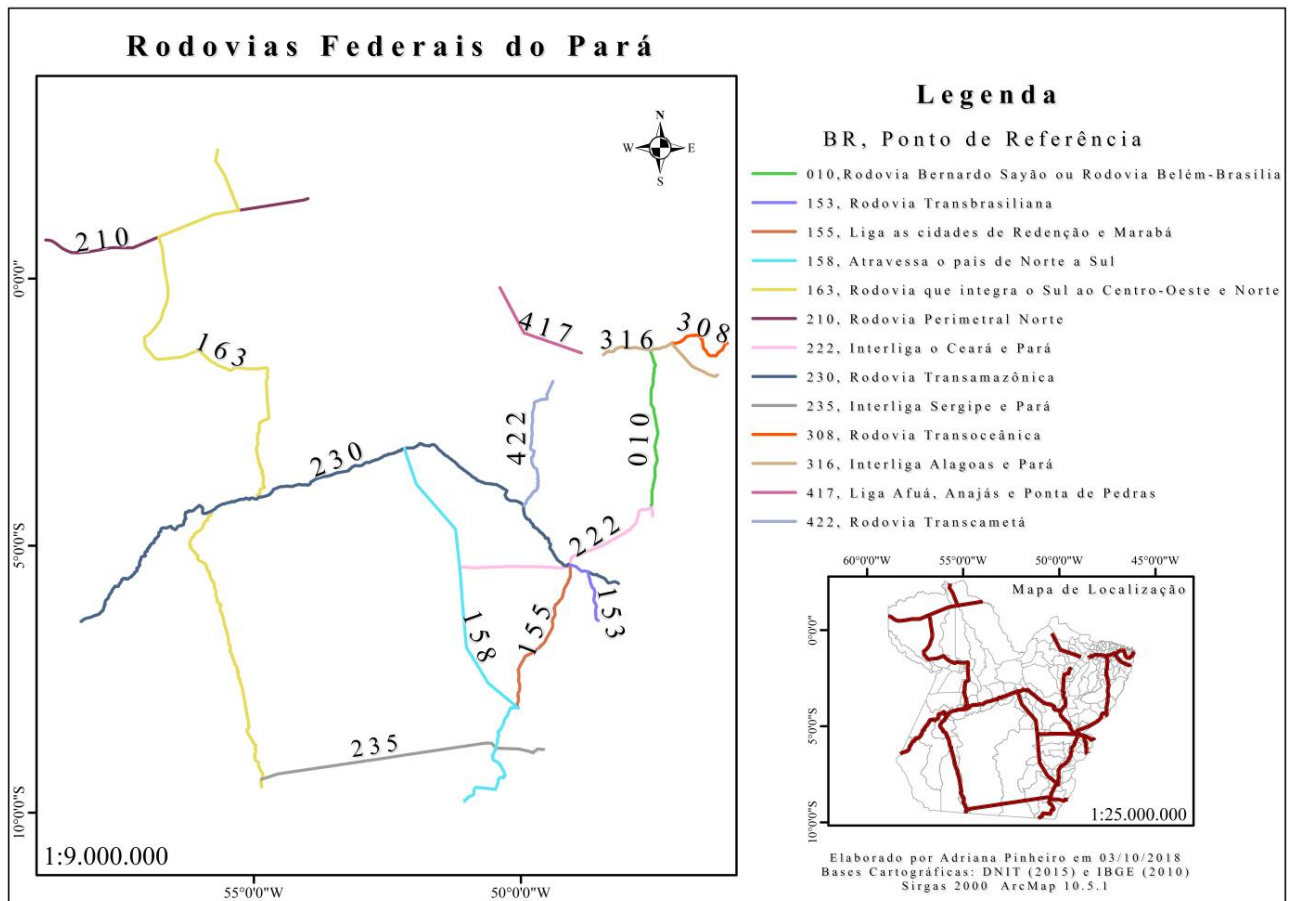
(Continua)

(Continuação)

BR	Extensão Total (Km)
163	5.317,30
210	1510,6
222	1.849,70
230	5.662,60
235	2.646,30
308	636,30
<b>316</b>	<b>2.067,60</b>
417	235
422	336

Fonte: Ministério dos Transportes, Portos e Aviação Civil.

Figura 8 - Rodovias Federais do Pará



Fonte: Autora.

## 2.4 MINERAÇÃO DE DADOS NA ANÁLISE DE ACIDENTES RODOVIÁRIOS

A sociedade está sobrecarregada de dados e parece continuar aumentando. A eletrônica onipresente registra decisões, escolhas no supermercado, hábitos financeiros, idas e vindas. O

mundo da Web sobrecarrega indivíduos com muitas informações. À medida que o volume de dados aumenta, inexoravelmente, a proporção disso que as pessoas entendem diminui, de forma alarmante. Informações potencialmente úteis, raramente são explicitadas ou aproveitadas.

As pessoas vêm buscando padrões nos dados desde que a vida humana começou. Caçadores buscam padrões no comportamento de migração animal, os agricultores buscam padrões na cultura de crescimento, os políticos buscam padrões na opinião dos eleitores. O trabalho de um cientista é dar sentido aos dados para descobrir os padrões que governam como o mundo físico funciona e encapsular-los em teorias que podem ser usados para prever o que vai acontecer em novas situações. O trabalho do empreendedor é identificar oportunidades, isto é, padrões no comportamento que pode ser transformado em um negócio lucrativo e explorá-los (WITTEN, I.H; FRANK, E; HALL, M.A. 2011).

A mineração de dados moderna começou em meados da década de 1990, quando o poder de computação e o custo de seu processamento e armazenamento finalmente alcançou um nível que permitia as empresas fazê-la internamente, sem precisar recorrer a recursos de computação externos. Na mineração de dados, os dados são armazenados eletronicamente e a pesquisa é automatizada - ou pelo menos aumentada - por computador. Isso não é particularmente novo. O que é novo é o aumento impressionante de oportunidades para encontrar padrões nos dados.

A mineração de dados é a transformação de grandes quantidades de dados em padrões e regras significativas. Para Rud (2011) a mineração de dados pode ser definida como um processo automatizado de descoberta de novas informações, em grandes massas de dados, sendo um passo essencial no processo de descoberta de conhecimentos.

Pode envolver o uso de diversas técnicas, tarefas e algoritmos. Basicamente o objetivo da mineração de dados é criar um modelo que possa melhorar o modo pelo qual lê-se e interpreta-se dados existentes e futuros.

Dias (2001) apresenta as tarefas e técnicas de mineração de dados de forma resumida, conforme demonstrado no quadro abaixo.

**Tabela 2** – Tarefas e Técnicas de Mineração de Dados.

Tarefa	Técnica	Descrição	Exemplos
<b>Classificação</b>	Árvores de Decisão; Algoritmos Genéticos; Redes Neurais Artificiais.	Constroi um modelo de algum tipo que possa ser aplicado a dados não classificados, a fim de categorizá-los em classes.	<ul style="list-style-type: none"> <li>• Classificar pedidos de crédito;</li> <li>• Esclarecer pedidos de seguros fraudulentos;</li> <li>• Identificar a melhor forma de tratamento de um paciente.</li> </ul>

(Continua)

(Continuação)

Tarefa	Técnica	Descrição	Exemplos
<b>Estimativa (ou Regressão)</b>	Árvores de Decisão.	Usada para definir um valor para alguma variável contínua desconhecida.	<ul style="list-style-type: none"> <li>• Estimar o número de filhos ou a renda total de uma família;</li> <li>• Estimar o valor, em tempo de vida, de um cliente;</li> <li>• Estimar a probabilidade de que um paciente morrerá, baseando-se nos resultados de diagnósticos médicos;</li> <li>• Prever a demanda de um consumidor para um novo produto.</li> </ul>
<b>Associação</b>	Descoberta de Regras de Associação.	Usada para determinar quais itens tendem a ocorrer juntos) em uma mesma transação.	<ul style="list-style-type: none"> <li>• Determinar quais produtos costumam ser colocados juntos em um carrinho de supermercado.</li> </ul>
<b>Sumarização</b>		Envolve métodos para encontrar uma descrição compacta para um subconjunto de dados.	<ul style="list-style-type: none"> <li>• Tabular o significado e desvios padrão para todos os itens de dados;</li> <li>• Derivar regras de síntese.</li> </ul>
<b>Segmentação (ou Clustering)</b>	Raciocínio Baseado em Casos ou MBR ( <i>Memory-Based Reasoning</i> ); Algoritmos Genéticos; Redes Neurais Artificiais.	Processo de partição de uma população heterogênea em vários subgrupos ou grupos mais homogêneos.	<ul style="list-style-type: none"> <li>• Agrupar clientes por região do país;</li> <li>• Agrupar clientes com comportamento de compra similar;</li> <li>• Agrupar seções de usuários web para prever comportamento futuro.</li> </ul>

**Fonte:** Adaptado de Dias, 2011.

Há muitas técnicas de mineração de dados. Não há uma técnica de mineração de dados que resolva todos os problemas. Cada método serve para diferentes aplicações, cada um oferece vantagens e desvantagens. A familiaridade com as técnicas é necessária e facilita a escolha, de acordo com os problemas apresentados a nível de mineração de dados.

#### 2.4.1 Regras de Associação

A técnica de descoberta de regras de associação é baseada em aprendizado não supervisionado e dispensa a classificação prévia dos dados e a interferência do especialista nesta atividade. Seus resultados são facilmente compreendidos pelo ser humano, e representam relações empíricas entre os dados, sendo possível a avaliação dessas relações por meio de medidas especiais, como o suporte e confiança (AGRAWAL, 1994; HAN; KAMBER, 2001).

"Descoberta de regras de associação é o processo de analisar os relacionamentos existentes entre atributos de uma base de dados transacional, com o objetivo de encontrar associações ou correlações." (SILVA, 2016, p. 199).

Dentre as utilizações mais comuns está a área de compras, quando se procura saber quais produtos são comprados associados. Por exemplo, pais que compram fraldas para seus bebês e que acabam por levarem também cervejas nos finais de semana. Com essa informação, a rede de supermercado pode colocar cervejas próximo às fraldas (GUROVITZ, 2011).

O software Weka tem seis algoritmos de regras de associação, listados na Tabela 3. Para esse estudo foi utilizado somente o algoritmo Apriori. Ele começa com um suporte mínimo de 100% dos itens de dados e diminui isso em etapas de 5% até que haja pelo menos 10 regras com a confiança mínima exigida de 0,9 ou até que o suporte tenha atingido um limite inferior de 10%, o que ocorrer primeiro. (Esses valores padrão podem ser alterados.) O Apriori tem uma opção para limitar as regras encontradas àquelas que contêm apenas o valor de um único atributo na consequência da regra. Tais regras são chamadas de regras de associação de “classe” - isto é, regras de classificação. As regras de associação geralmente envolvem apenas atributos não numéricos (WITTEN, I.H; FRANK, E; HALL, M.A. 2011).

**Tabela 3** - Tipos de Regras de associação no software Weka.

Nome	Função
<b>Apriori</b>	Gerar regras de associação que atendem aos limites mínimos de suporte e confiança. A priori segue uma metodologia de geração e teste para encontrar conjuntos de itens frequentes, gerando conjuntos de itens candidatos sucessivamente mais longos a partir de conjuntos mais curtos que são conhecidos como frequentes.
<b>FilteredAssociator</b>	Executar um associador em dados filtrados
<b>FPGrowth</b>	Mineração de regra de associação usando árvores de padrão frequentes
<b>GeneralizedSequentialPatterns</b>	Encontrar grandes conjuntos de itens em dados sequenciais
<b>PredictiveApriori</b>	Encontrar regras de associação classificadas por precisão preditiva
<b>Tertius</b>	Descoberta da regra de associação ou regra guiada por confirmação

**Fonte:** Adaptado de WITTEN, I.H; FRANK, E; HALL, M.A, 2011.

FilteredAssociator permite que os dados sejam passados por um filtro antes de chegar a um associador. Tanto o filtro quanto o associador base são opções que o usuário pode configurar.

PGrowth pode operar em instâncias padrão ou esparsas. A maioria de suas opções são as mesmas que para Apriori. Ele encontra o número solicitado de regras da mesma maneira - diminuindo iterativamente o suporte mínimo. Opcionalmente, o usuário pode ter FPGrowth para localizar todas as regras que atendem ao limite inferior para o suporte mínimo e o valor mínimo definido para a métrica de classificação (confiança, aumento, alavancagem ou convicção).

GeneralizedSequentialPatterns implementa uma versão do algoritmo GSP de Srikant e Agrawal (1996) para encontrar grandes conjuntos de itens que ocorrem sequencialmente ao longo do tempo. Os dados de entrada devem conter um atributo nominal especial que agrupe transações (instâncias) em tempo.

PredictiveApriori combina confiança e suporte em uma única medida de precisão preditiva (Scheffer, 2001) e encontra as melhores  $n$  regras de associação em ordem. Internamente, o algoritmo aumenta sucessivamente o limiar de suporte, porque o valor da precisão preditiva depende dele.

Tertius encontra regras de acordo com uma medida de confirmação (Flach e Lachiche, 1999), buscando regras com múltiplas condições no consequente, como a priori - mas diferindo em que essas condições são juntas ou não.

## **2.5 TRABALHOS CORRELATOS**

São poucas os trabalhos que tratam quantitativamente (seja em termos de abordagens estatísticas ou heurísticas) dados de acidentes de trânsito na rodovia BR-316. Em nossa pesquisa selecionamos o trabalho de Andrade et al (2011) que utilizou a técnica estatística de análise exploratória de dados para um estudo voltado a casos de acidentes de trânsito fatais, com base em fatos e dados.

O trabalho de Andrade et al (2011) tinha como finalidade gerar informações sobre os acidentes fatais de trânsito, ocorridos na Rodovia Federal BR-316, em 2008. Verificou-se que a maioria dos acidentes que acontecem ao longo da rodovia federal BR-316 ocorre nos quilômetros iniciais, em pista de mão dupla e de traçado reto, sendo que as vítimas na maior parte dos casos são os próprios condutores de veículos e os acidentes fatais acometem, em sua maioria, homens na faixa etária de 25 a 29 anos.

Já Reis (2014) identificou os fatores de contribuição de acidentes rodoviários, no período de 2008 a 2012, em segmentos críticos da rodovia BR-381. Através do software livre Weka e sua implementação do algoritmo Apriori, foram descobertas associações que

permitiram encontrar os principais segmentos críticos em pista simples e em pista dupla, bem como os padrões de associações para condutores do sexo masculino e feminino.

Por fim, Cruz e Santos (2018) utilizaram a ferramenta Weka para descobrir conhecimentos de acidentes de trânsito no município de Goiânia, no período de 2011 a 2017, a partir do banco da Secretaria de Segurança Pública do Estado de Goiás. Por meio de técnicas de mineração de dados, novos conhecimentos foram descobertos sobre as causas dos acidentes de trânsito em Goiânia, os períodos em que mais ocorrem acidentes, avaliação do número de vítimas, bem como a faixa etária dos envolvidos. No software Weka foram encontradas associações pelo algoritmo Apriori que podem ser utilizadas nas ações de mitigação dos acidentes de trânsito. Com os dados, tornou-se possível, também, a análise de vários gráficos gerados, de onde foram elencadas várias informações sobre o tema ora em análise. Também foi possibilitado a verificação de resultados de alguns algoritmos diferentes do Apriori, como o EM de clusterização. Os resultados não se mostraram tão eficazes para o objetivo proposto quanto resultados do algoritmo Apriori.

### 3. MATERIAS E MÉTODOS

Para Silva e Menezes (2005), a pesquisa quantitativa considera tudo que pode ser quantificável, ou seja, traduzir números, opiniões e informações para classificá-los e analisá-los. Este tipo de pesquisa requer o uso de algum recurso ou técnica estatística: percentagem, média, desviopadrão, análise de regressão, etc.

De acordo com Vergara (2007), os tipos de pesquisa podem ser definidos por dois critérios básicos: quanto aos fins e quanto aos meios. Quanto aos fins, a pesquisa descritiva expõe as características de um fenômeno ou de uma determinada população. Ela pode também correlacionar variáveis e definir a sua natureza, porém não tem o compromisso de explicar os fenômenos que descreve, embora sirva de base para tal explicação.

Quanto aos meios a pesquisa, pode ser classificada como pesquisa experimental - investigação empírica na qual o pesquisador controla e manipula variáveis independentes e observa as variações que tais manipulações produzem em variáveis dependentes. Desta forma, a metodologia utilizada no presente trabalho é apresentada na Tabela 4.

**Tabela 4** – Metodologia utilizada na Pesquisa.

<b>Metodologia de Pesquisa</b>	
<b>Tipo de Pesquisa</b>	Descritiva experimental
<b>Abordagem metodológica</b>	Quantitativa
<b>Análise dos dados</b>	Quantitativo envolvendo a análise por meio de mineração de dados para a descoberta de conhecimento da rodovia BR-316 com o uso do algoritmo Apriori do software livre WEKA.
<b>Critério de seleção dos dados</b>	Período restringido entre janeiro de 2007 a setembro de 2018 do banco de dados online da PRF.

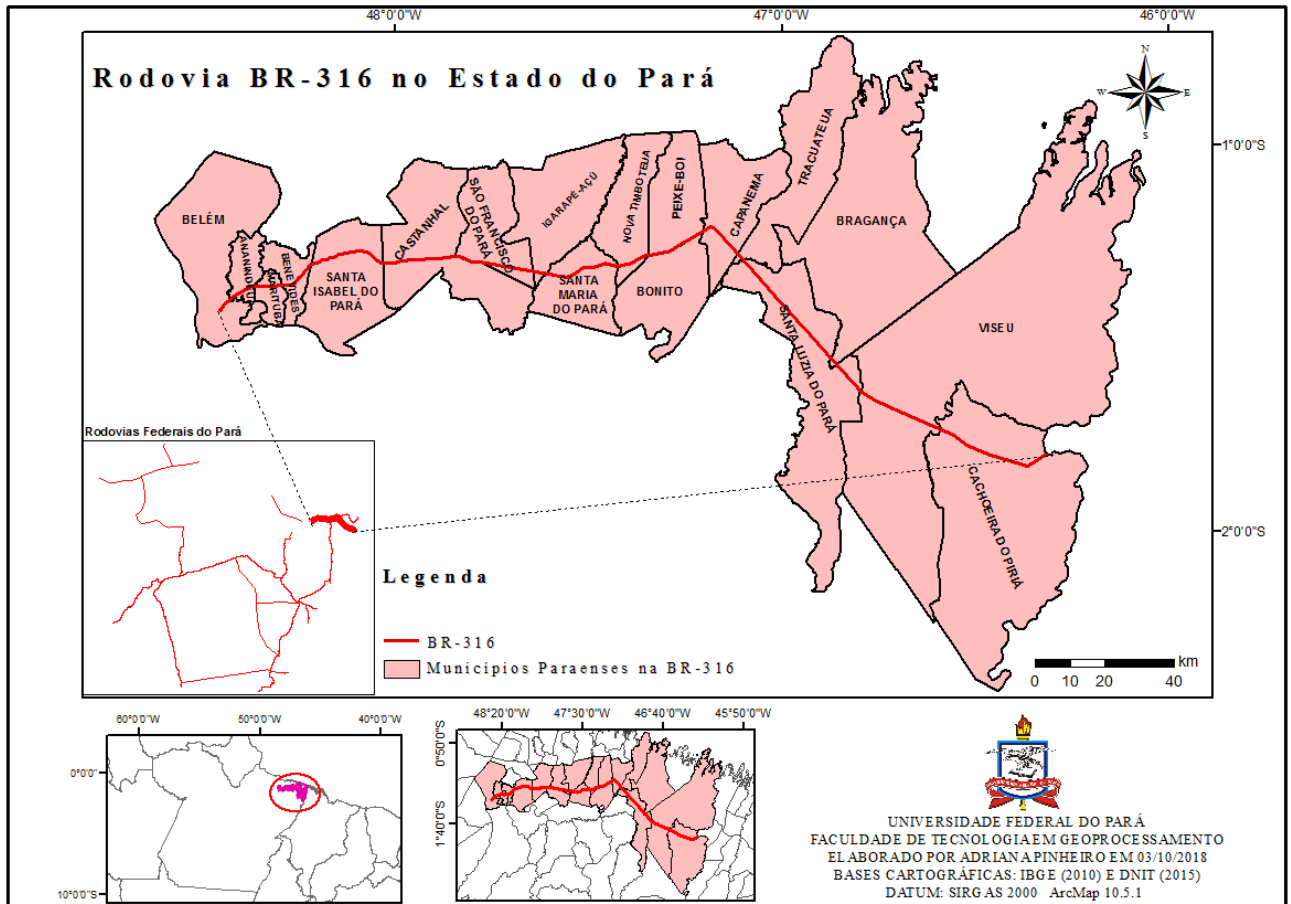
**Fonte:** Autora.

Para esse estudo foram usadas algumas ferramentas importantes. Além da principal ferramenta Weka 3.8, foi usado o software ArcGis 10.5.1, licenciado para o Laboratório de Geoprocessamento da UFPA – Campus Ananindeua, bem como as bases cartográficas de rodovias federais do DNIT e municípios do IBGE para fazer os mapas presentes neste trabalho, tal como o bloco de notas do Windows 10, e o Excel 2016.

### 3.1 Banco de dados da polícia rodoviária federal

Inicialmente foram identificadas as cidades que ligam a BR-316 (Figura 9) no estado do Pará para um melhor entendimento da área de estudo.

**Figura 9** - Rodovia BR-316 no Estado do Pará



**Fonte:** Autora.

Posteriormente buscou-se junto ao Grupo de Educação para o Trânsito (GETRAN), da Superintendência da Polícia Rodoviária Federal (PRF) no Pará, as informações sobre os acidentes de trânsito, bem como o entendimento sobre o gerenciamento de registro de acidentes de trânsito. Em seguida, foi feito *download* da base de dados disponibilizadas de acidentes da BR-316 a partir do Boletim de Acidente de Trânsito (BAT) da base de dados do Departamento da Polícia Rodoviária Federal (Datatran/DPRF) em formato CSV disponível no site da Polícia Rodoviária Federal (Figura 10).

**Figura 10** - Página do Departamento da Polícia Rodoviária Federal (Datatran/DPRF)



Fonte: Página do Departamento de Polícia Rodoviária Federal, 2018.

Foi feito o *download* de 12 planilhas zipadas agrupadas por ocorrência sobre dados de acidentes de trânsito e com uma média de 3 MB (*MegaBytes*) cada uma.

### 3.2 SOFTWARE WEKA

O software WEKA (*Waikato Environment for Knowledge Analysis*) é uma ferramenta mantida pelo grupo de aprendizagem de máquina da Universidade de Waikato na Nova Zelândia.

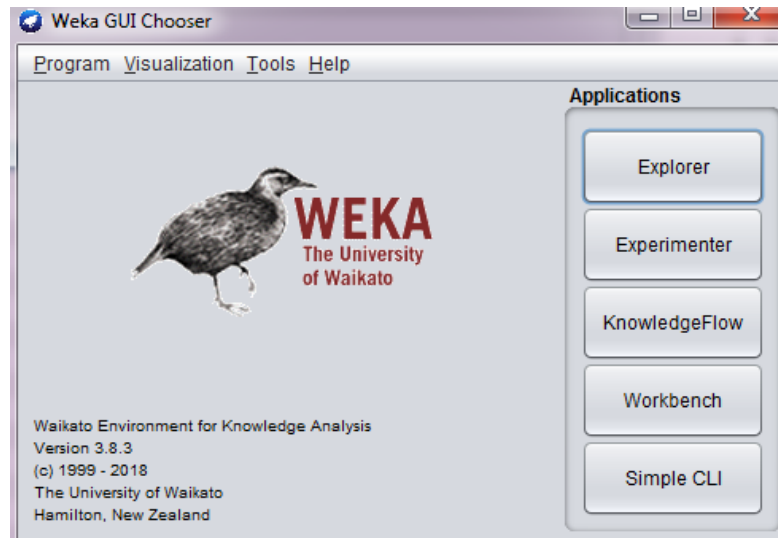
Foi implementado pela primeira vez em sua forma moderna em 1997. Ele usa a licença GNU<sup>2</sup>. O software foi escrito na linguagem Java™ e contém uma Interface gráfica do usuário - GUI (Graphical User Interface) para interagir com arquivos de dados e produzir resultados visuais (Tabelas e curvas).

Weka também tem uma API (*Application Programming Interface*), assim é possível incorporar o software, como qualquer outra biblioteca, a seus próprios aplicativos para fazer tarefas de mineração de dados automatizadas no lado do servidor. Possui uma coleção de algoritmos que podem ser utilizados tanto comercialmente quanto em ambiente acadêmico. Estão presentes ferramentas para pré-processamento e visualização dos dados, classificação, regressão, agrupamento e regras de associação.

O Weka realiza a leitura de bases de dados de duas fontes: arquivos com a extensão ARFF e também por meio de conexão direta com bancos de dados. São disponibilizados arquivos de configuração para o Access, SQL Server, Oracle, PostgreSQL e HSQLDB (com JDBC), bem como acesso via ODBC para os outros SGDBs.

<sup>2</sup> GNU: A Licença Pública Geral GNU é uma licença copyleft gratuita para software.

**Figura 11** - Tela Inicial do Weka



### 3.3 PROCEDIMENTOS METODOLÓGICOS

Foram realizados diversos procedimentos para adequar os arquivos das planilhas da Polícia Rodoviária Federal, além de configurações no software Weka de modo que pudéssemos executar os algoritmos de associação de maneira correta.

#### 3.3.1 OBTENÇÃO DE DADOS

Os dados sobre acidentes de trânsito são consolidados e produzidos a partir de agentes públicos de trânsito, sendo eles em três esferas, municipal, estadual e, federal, portanto são os primeiros órgãos que devem ser procurados para a coleta de dados.

No estudo em questão foram inseridos somente dados disponíveis do site da PRF, ou seja, foram usados dados registrados por agente público federal, haja vista que a BR-316 é uma rodovia federal. As planilhas continham informações que encontram-se no Apêndice A.

Na base de dados do Departamento da Polícia Rodoviária Federal foram baixadas as planilhas em formato CSV dos anos de 2007 a Setembro de 2018. As planilhas contêm informações de acidentes de trânsito de todas as rodovias federais do Brasil, mas foram filtradas informações referentes ao estado do Pará, e mais especificamente da BR-316. A Tabela 5 mostra a quantidade de registros de acidentes relativos as rodovias federais de todo o Brasil, de todo o Pará, e da BR-316, respectivamente.

**Tabela 5** – Quantidade de Registros de Acidentes de Trânsito Processados.

Ano	Quantidade		
	Brasil	Pará	BR-316
2007	127675	3128	2225
2008	141043	3299	2360
2009	158646	3413	2507
2010	183469	4291	3262
2011	192326	4577	3418
2012	184569	3842	2572
2013	186749	3601	2293
2014	169202	3232	2123
2015	122162	2318	1344
2016	96364	1733	943
2017	89518	1503	787
Setembro/2018	51983	1014	496

**Fonte:** Autora

Na Tabela é possível verificar que há uma diminuição de acidentes de trânsito nos últimos três anos. Isso se deve ao fato do intenso trabalho da PRF que trabalha com prevenção de acidentes de trânsito, bem como aplicação de leis mais rigorosas, como a Lei Seca e aos fechamentos de alguns retornos da BR-316 no Pará.

### 3.3.2 PRÉ-PROCESSAMENTO DOS DADOS

Depois de ter feito o *download* das 12 planilhas, foi realizado um pré-processamento no Excel e posteriormente no Weka.

No Excel, usando o filtro, foram retirados dados de acidentes dos outros estados, e das BR's que não fizeram parte do estudo, ou seja, foram deixados somente informações referentes ao estado do Pará e da BR-316.

Houve a retirada de informações inconsistentes como um registro de acidente no município de Ourém na planilha do ano de 2010, condição meteorológica como granito, campos sem informações e de campos como coordenadas geográficas, delegacia e regional, sendo que essas três últimas informações só continham a partir do ano de 2017.

Também foi retirado os campos BR e UF, pois os dados eram iguais para todos, e no campo horário foi retirado os segundos, deixando somente a hora e minuto.

Em seguida, houve uma substituição dos campos onde havia espaço entre as palavras, substituído por *underline* para evitar possíveis erros; e, por fim no Excel, o campo *data\_inversa* foi extraído, de modo que ficasse um campo somente com o dia, outro com o mês e outro com o ano. Para esse processo foram usadas as funções *dia*, *mês* e *ano*, respectivamente.

- Função Dia: =DIA(núm\_série)
- Função Mês: =MÊS(núm\_série)
- Função Ano: =ANO(núm\_série)

**Figura 12 - Exemplo da extração**

	A	B	C	D
1	data_inversa	DIA	MESES	ANO
2	01/01/2015	1	1	2015

Depois de ser pré-processado no excel, ficaram 24 variáveis como mostra o dicionário de dados a seguir (Tabela 6).

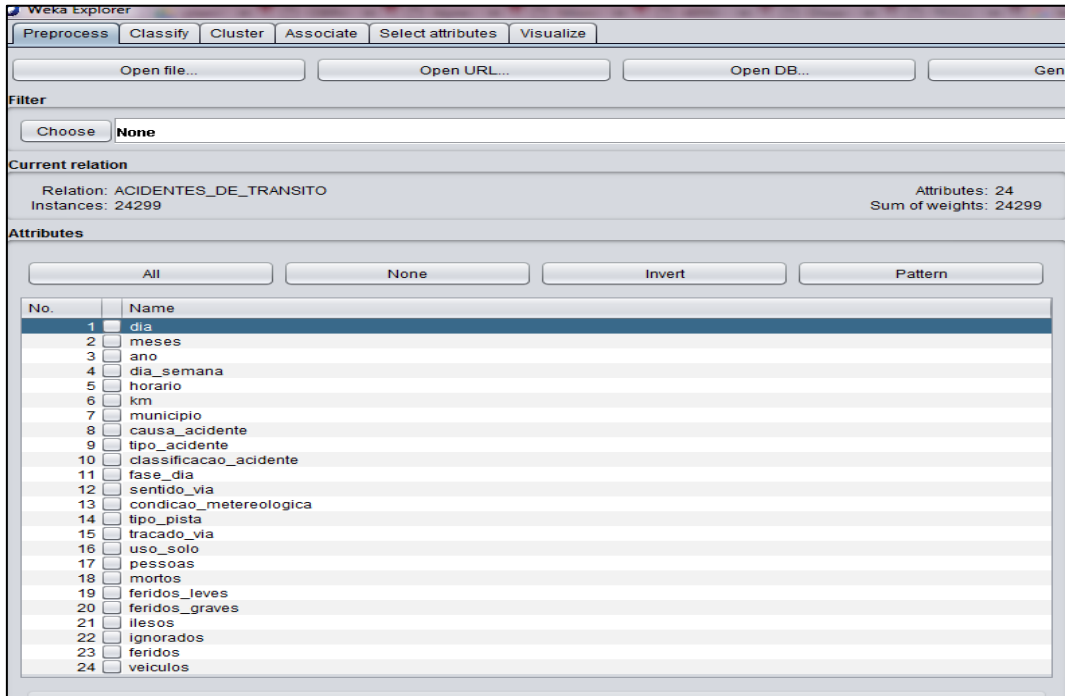
**Tabela 6 - Dicionário de Dados**

Variável	Descrição	Tipo de Dado
Dia	Dia da ocorrência	Nominal
Meses	Mês do ocorrência	Nominal
Ano	Ano do ocorrência	Nominal
dia_semana	Dia da semana do ocorrência	Nominal
Horário	Horário da ocorrência	Data
Km	Km da BR-316	Numérico
municipio	Município do Pará	Nominal
causa_acidente	Causa do acidente	Nominal
tipo_acidente	Tipo do acidente	Nominal
classificacao_acidente	Classificação do acidente	Nominal
fase_dia	Fase do dia da ocorrência	Nominal
sentido_via	Sentido da via	Nominal
condicao_metereologica	Condição metereológica da ocorrência	Nominal
tipo_pista	Tipo de Pista	Nominal
tracado_via	Traçado da Via	Nominal
uso_solo	Uso do solo (Urbano ou Rural)	Nominal
Pessoas	Pessoas envolvidas	Numérico
Mortos	Mortos no acidente	Numérico
feridos_leves	Feridos leves no acidente	Numérico
feridos_graves	Feridos graves no acidente	Numérico
Ilesos	Ilesos no acidente	Numérico
ignorados	Pessoas ignoradas	Numérico
Feridos	Feridos no acidente	Numérico
veiculos	Veículos envolvidos no acidente	Numérico

**Fonte:** Autora

Após os dados serem transformados em arquivo ARFF e importados no ambiente Weka, alguns atributos foram pré-processados na janela de pré-processamento (Figura 13).

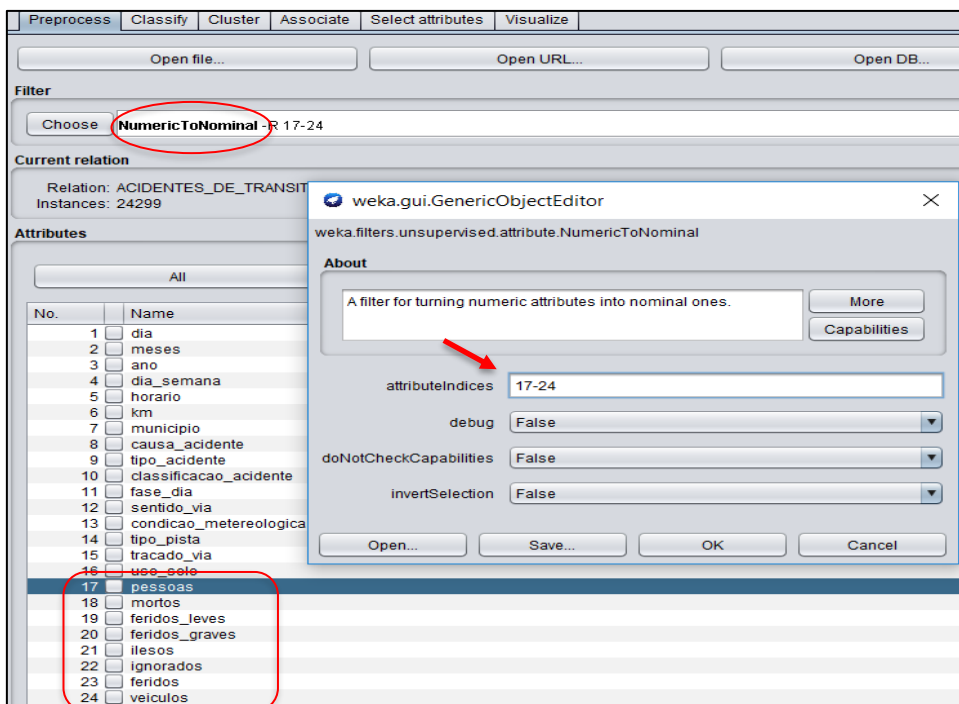
**Figura 13** - Janela de pré-processamento no Weka



Fonte: Autora

Houve a conversão (Figura 14) dos atributos 17 a 24 de numérico para nominal, distribuindo assim, todos os valores.

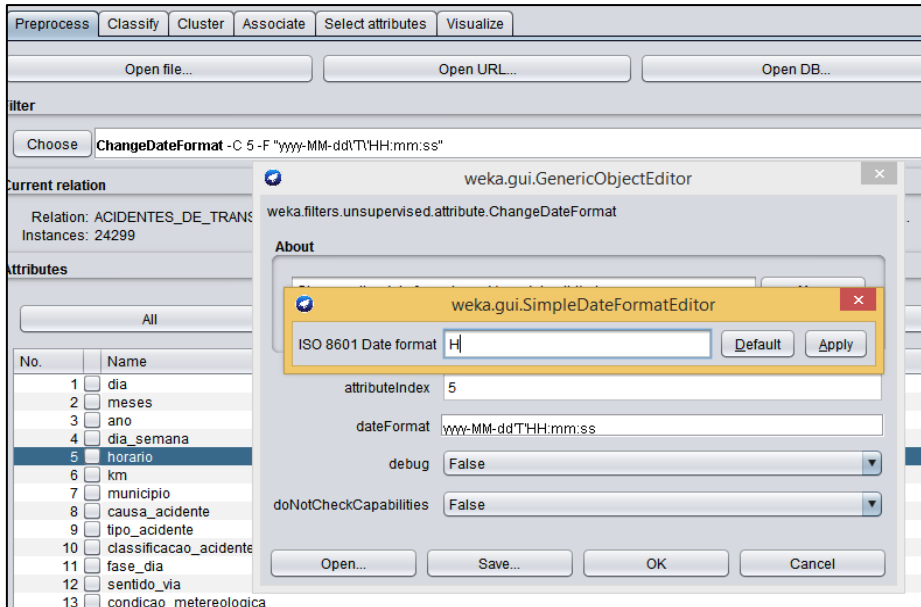
**Figura 14** - Aplicação do filtro NumericToNominal



Fonte: Autora

Também foi realizada no Weka a conversão do atributo horário pelo filtro *changedateformat* de forma que ficasse somente a hora, excluindo os minutos (Figura 15).

**Figura 15** - Conversão do Atributo Horário



Fonte: Autora

Em seguida o atributo horário foi convertido para nominal e logo após foi usado o filtro *mergemanyvalues* para unir os horários, como mostra a Figura 16.

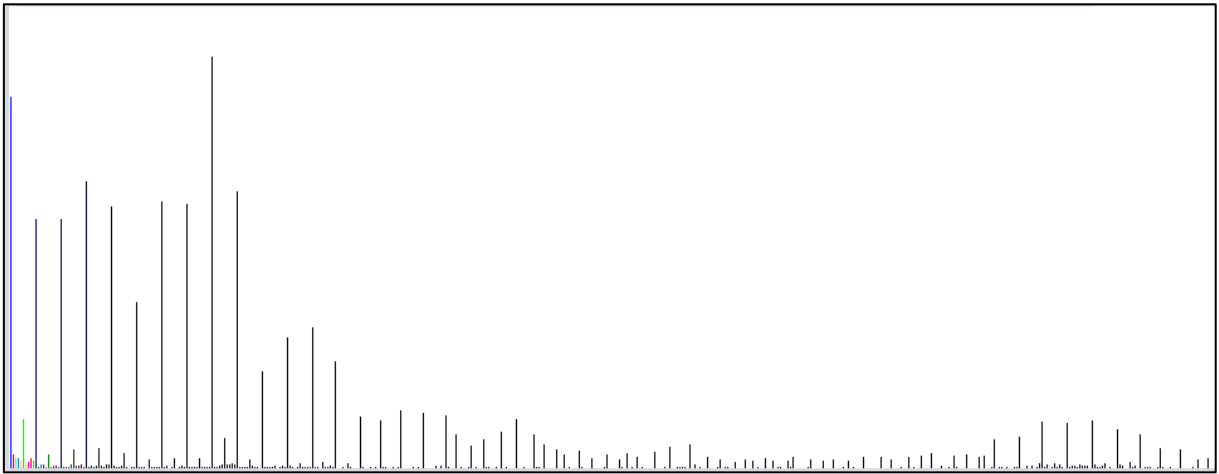
**Figura 16** - Distribuição do Horário após a aplicação do Filtro *mergemanyvalues*

Name: horario		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 4	
No.	Label	Count	Weight
1	00:00-05:00	1443	1443.0
2	06:00-11:00	7809	7809.0
3	12:00-17:00	8592	8592.0
4	18:00-23:00	6455	6455.0

Fonte: Autora

Havia 880 valores diferentes no atributo km, ou seja, continha uma variabilidade muito grande (Figura 17). Por isso, foi usado o filtro *Discretize*, o qual o Weka encontrou um padrão e fez uma distribuição de frequência com 10 intervalos, e em seguida transformou para nominal em um processo automatizado. Foi usado o filtro *rename nominal values* para renomear o nome dos intervalos.

**Figura 17** - Distribuição do atributo Km antes de ser usado o filtro *Discretize*.



### 3.3.3 TRANSFORMAÇÃO DOS DADOS

Para que pudesse ser aberto pela ferramenta Weka, o arquivo CSV foi transformado para o formato ARFF. Um arquivo ARFF (Formato de Arquivo de Relação de Atributos) é um arquivo de texto ASCII que descreve uma lista de instâncias que compartilham um conjunto de atributos. Os arquivos ARFF foram desenvolvidos pelo Machine Learning Project no Departamento de Ciência da Computação da Universidade de Waikato para uso com o software de aprendizado de máquina Weka (The University of Waikato, 2018).

Depois da conversão o arquivo “acidentes\_de\_transito” pode ser visualizado a partir de um editor de texto simples, neste caso foi usado o bloco de notas do Windows 10. Verificou-se que mesmo salvando em CSV, o arquivo foi separado por ponto e vírgula. Com isso, no bloco de notas, os dados foram substituído por vírgula, seguindo o formato do arquivo ARFF, e por fim foi montado um cabeçalho e salvo como acidentes\_de\_transito.ARFF.

O Apêndice B contém o cabeçalho feito no bloco de notas, ou seja, a estrutura do arquivo da Figura 18 que foi usada para ser executada no Weka.

**Figura 18 - Estrutura final do arquivo ARFF**

```

RELATION ACIDENTES_DE_TRANSITO
ATTRIBUTE dia {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31}
ATTRIBUTE meses {janeiro, fevereiro, marco, abril, maio, junho, julho, agosto, setembro, outubro, novembro, dezembro}
ATTRIBUTE ano {2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018}
ATTRIBUTE dia_semana {segunda, terca, quarta, quinta, sexta, sabado, domingo}
ATTRIBUTE horario date "HH:mm"
ATTRIBUTE km numeric
ATTRIBUTE municipio {ananindeua, belem, benevides, cachoeira_do_piria, capanema, castanhal, Igarape_açu, marituba, nova_timboteua, sant
ATTRIBUTE causa_acidente {animais_na_pista, defeito_na_via, desobediencia_a_sinalizacao, dormindo, falta_de_atencao, ingestao_de_alcool
ATTRIBUTE tipo_acidente {atropelamento_de_animal, atropelamento_de_pessoa, capotamento, colisao_com_bicicleta, colisao_com_objeto_fixo,
ATTRIBUTE classificacao_acidente {com_vitimas_fatais, com_vitimas_feridas, ignorado, sem_vitimas}
ATTRIBUTE fase_dia {amanhecer, anoitecer, pleno_dia, plena_noite}
ATTRIBUTE sentido_via {crescente, decrescente}
ATTRIBUTE condicao_meteorologica {ceu_claro, chuva, ignorado, nevoeiro_neblina, nublado, sol, vento}
ATTRIBUTE tipo_pista {dupla, multipla, simples}
ATTRIBUTE tracado_via {cruzamento, curva, reta, desvio_temporario, intersecao_de_vias, nao_informado, ponte, viaduto, retorno_regulamer
ATTRIBUTE uso_solo {rural, urbano}
ATTRIBUTE pessoas numeric
ATTRIBUTE mortos numeric
ATTRIBUTE feridos_leves numeric
ATTRIBUTE feridos_graves numeric
ATTRIBUTE ileso numeric
ATTRIBUTE ignorados numeric
ATTRIBUTE feridos numeric
ATTRIBUTE veiculos numeric

#data
1,janeiro,2016,sexta,08:15,78.8,sao_francisco_do_para,desobediencia_a_sinalizacao,colisao_lateral,com_vitimas_feridas,pleno_dia,crescent
2,janeiro,2016,domingo,10:30,111.5,santa_maria_do_para,nao_guardar_distancia_de_seguranca,colisao_traseira,sem_vitimas,pleno_dia,decresc
3,janeiro,2016,domingo,12:30,111.5,santa_maria_do_para,nao_guardar_distancia_de_seguranca,colisao_traseira,sem_vitimas,pleno_dia,decresc
4,janeiro,2016,domingo,10:30,217,santa_luzia_do_para,outros,colisao_com_objeto_movel,com_vitimas_feridas,pleno_dia,decrecente,ceu_claro,dupla
5,janeiro,2016,domingo,12:40,11,marituba,desobediencia_a_sinalizacao,colisao_lateral,com_vitimas_feridas,pleno_dia,decrecente,sol,dupla
6,janeiro,2016,segunda,11:40,9,marituba,falta_de_atencao,colisao_traseira,sem_vitimas,pleno_dia,decrecente,ceu_claro,dupla,reta,urbano,2
7,janeiro,2016,segunda,17:15,11,marituba,falta_de_atencao,colisao_traseira,com_vitimas_feridas,anoitecer,decrecente,ceu_claro,dupla,ret
8,janeiro,2016,segunda,22:00,88,santa_maria_do_para,outros,colisao_transversal,com_vitimas_feridas,plena_noite,decrecente,ceu_claro,sim
9,janeiro,2016,quarta,16:20,2,ananindeua,nao_guardar_distancia_de_seguranca,colisao_traseira,sem_vitimas,pleno_dia,decrecente,ceu_claro,dupla
10,janeiro,2016,quarta,19:20,0,belem,nao_guardar_distancia_de_seguranca,colisao_lateral,sem_vitimas,plena_noite,crescente,ceu_claro,dupla
11,janeiro,2016,segunda,01:00,46,santa_isabel_do_para,ingestao_de_alcool,colisao_traseira,sem_vitimas,plena_noite,decrecente,nublado,dupla
12,janeiro,2016,quinta,17:30,11,ananindeua,defeito_na_via,queda_de_motocicleta_bicicleta_veiculo,com_vitimas_feridas,anoitecer,decrecent
13,janeiro,2016,sabado,09:45,8,ananindeua,falta_de_atencao,colisao_transversal,sem_vitimas,pleno_dia,crescente,ceu_claro,dupla,reta,urban
14,janeiro,2016,sabado,11:45,203,santa_luzia_do_para,falta_de_atencao,colisao_frontal,com_vitimas_fatais,pleno_dia,decrecente,sol,simple
15,janeiro,2016,sabado,14:30,30,santa_isabel_do_para,outros,colisao_traseira,sem_vitimas,pleno_dia,crescente,chuva,dupla,curva,rural,3,0
16,janeiro,2016,sabado,18:00,1,belem,nao_guardar_distancia_de_seguranca,colisao_traseira,sem_vitimas,anoitecer,decrecente,nublado,dupla,
17,janeiro,2016,sabado,20:30,3,ananindeua,falta_de_atencao,colisao_transversal,sem_vitimas,plena_noite,crescente,ceu_claro,simples,reta,u
18,janeiro,2016,sabado,18:30,8,ananindeua,ingestao_de_alcool,colisao_traseira,com_vitimas_feridas,anoitecer,crescente,ceu_claro,dupla,ret
19,janeiro,2016,sabado,23:00,3,ananindeua,falta_de_atencao,colisao_lateral,com_vitimas_feridas,plena_noite,decrecente,ceu_claro,dupla,ret
20,janeiro,2016,segunda,02:15,8,ananindeua,outros,colisao_com_objeto_fixo,com_vitimas_feridas,plena_noite,crescente,ignorado,dupla,reta,
21,janeiro,2016,segunda,09:45,19,benevides,nao_guardar_distancia_de_seguranca,colisao_traseira,com_vitimas_feridas,pleno_dia,crescente,c
22,janeiro,2016,segunda,10:30,54,castanhal,falta_de_atencao,colisao_traseira,com_vitimas_feridas,pleno_dia,crescente,ceu_claro,dupla,ret
23,janeiro,2016,segunda,13:00,4,ananindeua,outros,colisao_traseira,com_vitimas_feridas,pleno_dia,decrecente,ceu_claro,dupla,reta,urban
24,janeiro,2016,quarta,13:05,6,ananindeua,ultrapassagem_indevida,colisao_lateral,sem_vitimas,pleno_dia,decrecente,sol,dupla,reta,rural,3
25,janeiro,2016,terca,08:00,7,ananindeua,falta_de_atencao,colisao_lateral,com_vitimas_feridas,pleno_dia,crescente,ceu_claro,dupla,reta,u
26,janeiro,2016,terca,07:57,10,marituba,nao_guardar_distancia_de_seguranca,colisao_traseira,sem_vitimas,pleno_dia,decrecente,ceu_claro,
27,janeiro,2016,domingo,06:00,6,ananindeua,falta_de_atencao,colisao_com_objeto_fixo,com_vitimas_feridas,amanhecer,decrecente,ceu_claro,
28,janeiro,2016,quinta,02:40,1,ananindeua,dormindo,colisao_com_objeto_fixo,com_vitimas_feridas,plena_noite,crescente,ceu_claro,dupla,ret
29,janeiro,2016,quinta,10:40,9,marituba,falta_de_atencao,colisao_lateral,sem_vitimas,pleno_dia,crescente,ceu_claro,dupla,reta,urbano,2,0
30,janeiro,2016,quinta,13:20,9,marituba,falta_de_atencao,colisao_lateral,sem_vitimas,amanhecer,decrecente,ceu_claro,dupla,reta,urbano,2
31,janeiro,2016,quinta,16:00,2,ananindeua,ingestao_de_alcool,colisao_transversal,com_vitimas_feridas,pleno_dia,decrecente,nublado,dupla

```

Fonte: Autora.

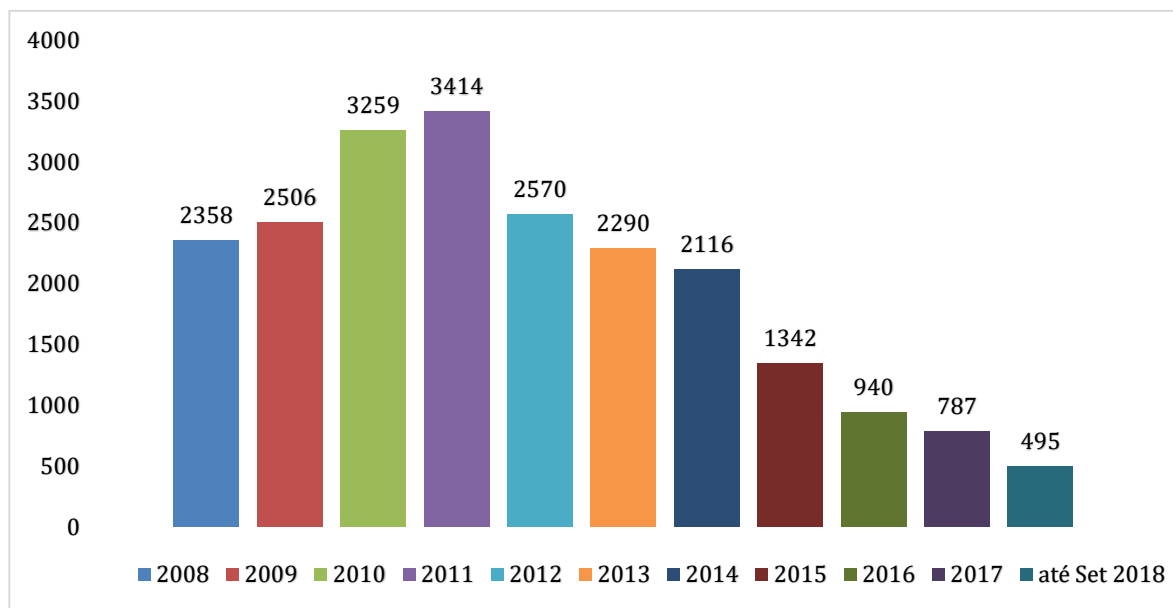
## 4. RESULTADOS E DISCUSSÕES

Neste capítulo, serão apresentados os resultados estatísticos dos principais atributos e a análise das regras mais relevantes que forneceram novos conhecimentos dos fatores de contribuição dos acidentes na BR-316, no período de janeiro de 2007 a setembro de 2018. A seção 4.1 apresenta a análise dos dados estatísticos gerais dos acidentes de trânsito da BR-316 no Pará no período em questão. A seção 4.2 apresenta os resultados encontrados envolvendo os atributos horário, km, município e causa do acidente. E por fim, a seção 4.3 apresenta os resultados envolvendo os atributos dia da semana, município, causa do acidente e tipo de acidente.

### 4.1 ANÁLISE DOS DADOS ESTATÍSTICOS GERAIS DOS ACIDENTES DE TRÂNSITO DA BR-316 NO PARÁ ENTRE JANEIRO DE 2007 A SETEMBRO DE 2018

De um total de 170.370 registros contidos no banco de dados da Polícia Rodoviária Federal, 35.951 (aproximadamente 22%) eram acidentes ocorridos em todo o estado do Pará, e 24.330 (aproximadamente 15%) eram acidentes de trânsito ocorridos somente na BR-316 no estado do Pará. A evolução em quantidade para cada ano pode ser observada na Figura 19.

**Figura 19** - Distribuição anual dos acidentes da BR-316 no estado do Pará

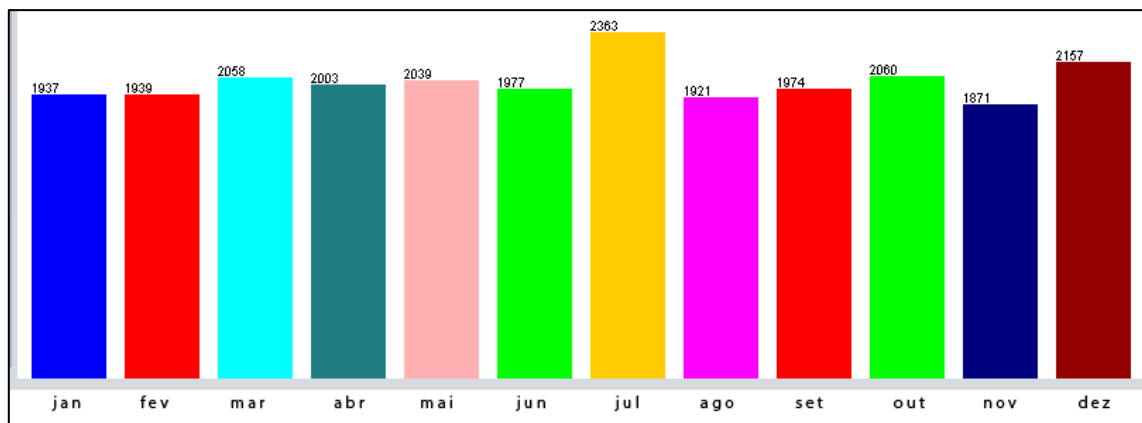


Fonte: Autora

Observa-se um grande aumento de acidentes entre 2009 e 2011. Um total de 908 acidentes a mais em 2011, quando comparado ao ano de 2009. A diminuição de acidentes nos últimos anos provavelmente se deve às aplicações de leis mais rígidas e ao fechamento de alguns retornos ao longo da BR-316 no Pará.

Se a análise for feita pelo mês (Figura 20) observa-se que os meses de Julho e Dezembro lideram o *ranking* com mais acidentes de trânsito. O Pará, em transporte viário, só tem a BR-316 como saída para cidades com praias, como Cametá, Salinópolis e Abaetetuba. A falta de rodovias alternativas pode intensificar o aumento de acidentes na BR-316.

**Figura 20** - Distribuição mensal dos Acidentes de Trânsito no estado do PA entre os anos de 2007 a setembro de 2018

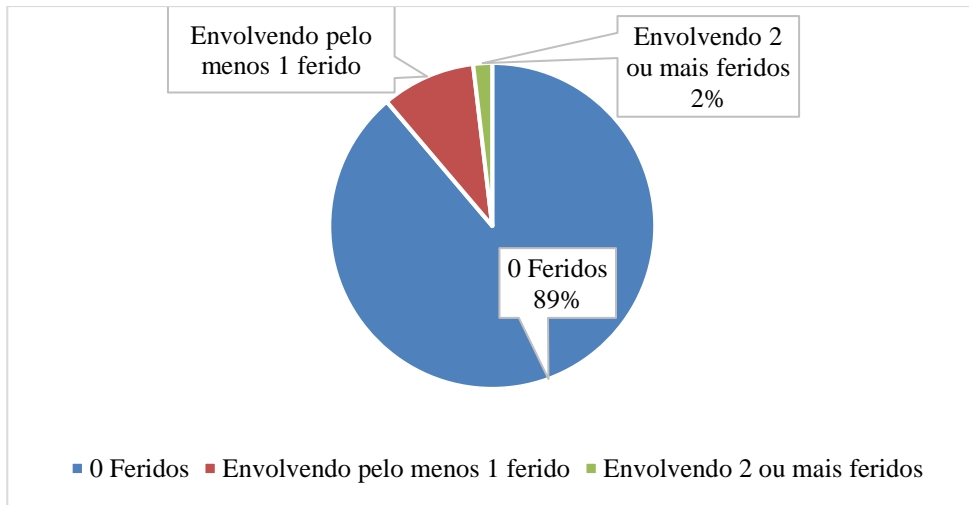


**Fonte:** Autora

A análise pela quantidade total de feridos graves (Figura 21), mostra que houve um total de 2.265 (9%) vítimas em estado grave com pelo menos uma pessoa.

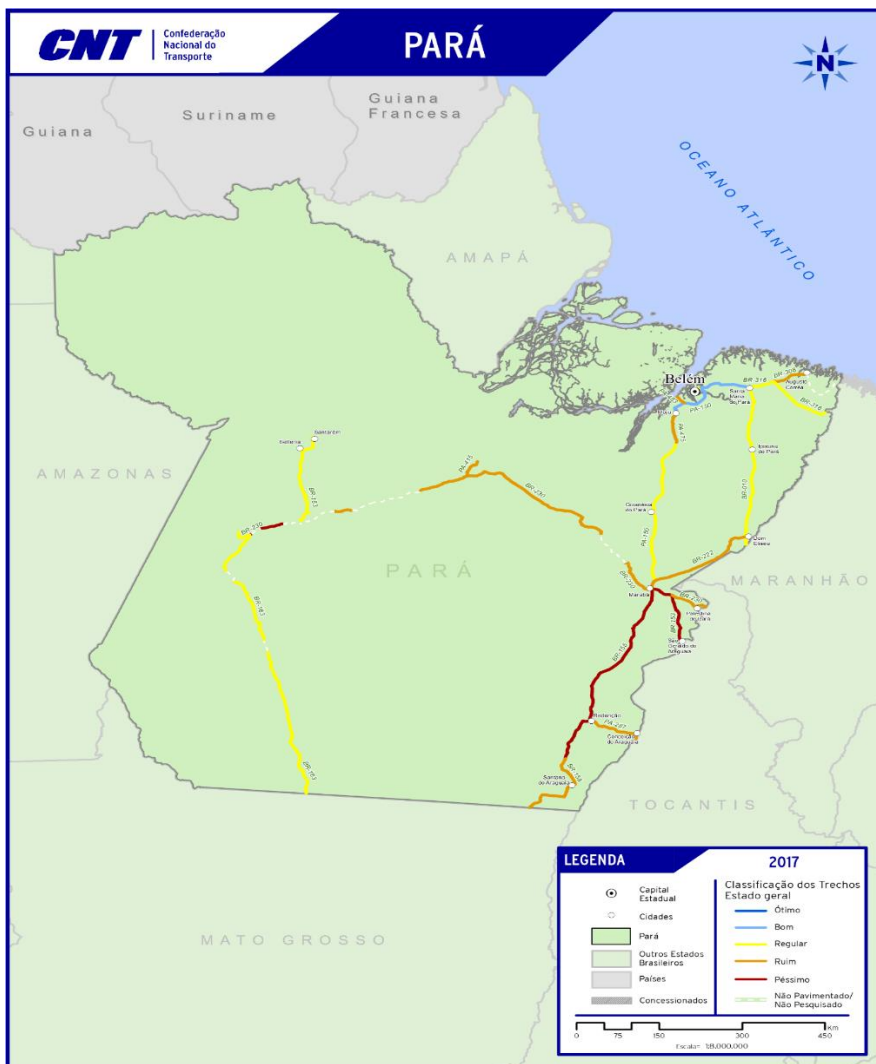
De acordo com o recente relatório de 2018 da situação global sobre segurança no trânsito feito pela OMS, a infra-estrutura rodoviária está fortemente ligada à causalidade de ferimentos graves e fatais em colisões de tráfego rodoviário. Outra pesquisa sobre rodovias feita pela Confederação Nacional do Transporte (2017) mostrou que a qualidade do trecho rodoviário da BR-316 entre Santa Maria do Pará e Cachoeira do Piriá, é regular (Figura 22). Isso representa que a rodovia precisa de melhorias ser tratada como prioridade pelos gestores do estado do PA.

**Figura 21** - Quantidade de ferido graves dos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018



Fonte: Autora

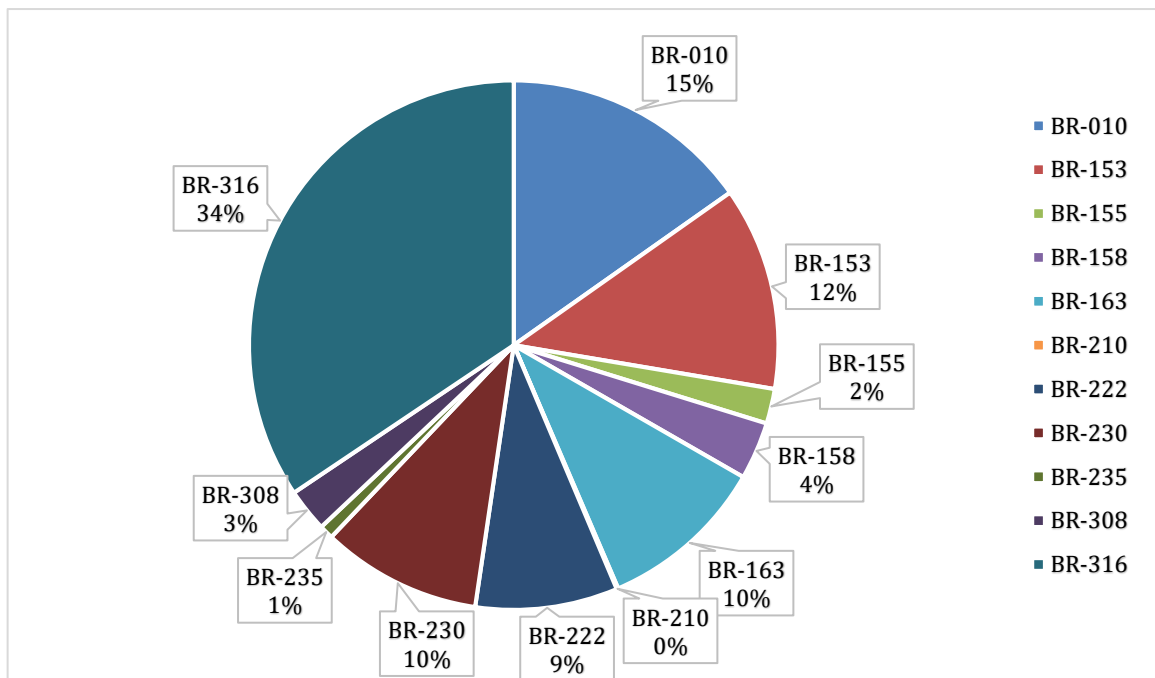
**Figura 22** - Classificação dos Trechos das Rodovias do Pará



Fonte: CNT, 2017.

A quantidade total de mortos somente da rodovia BR-316 no período equivalente entre Janeiro de 2007 a Setembro de 2018 resultou em 1.085 (Tabela 7). De todas as rodovias federais do Pará, não havia dados da BR-417, logo, foi realizada a análise somente com 11 rodovias. Para esta análise foram usados dados referentes a essas rodovias disponibilizados, também, na base de dados do (Datatran/DPRF). A distribuição total (Figura 23) de mortes abaixo, mostra que a BR-316 lidera no *ranking* com mais mortes de acidentes de trânsito em todo o estado.

**Figura 23** - Distribuição total de Mortes das Rodovias Federais do Pará entre Janeiro de 2007 a Setembro de 2018



**Fonte:** Autora

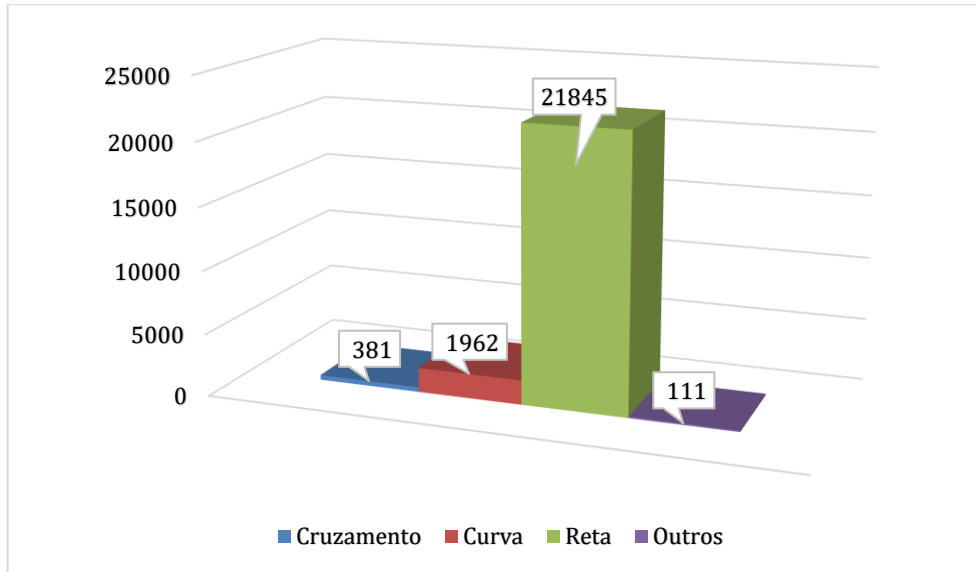
**Tabela 7** - Quantidade total de Mortes das Rodovias do Pará entre Janeiro de 2007 a Setembro de 2018

Rodovias	Quantidade Total de Mortes entre Janeiro de 2007 a Setembro de 2018
BR-010	480
BR-153	392
BR-155	67
BR-158	111
BR-163	323
BR-210	3
BR-222	274
BR-230	308
BR-235	27
BR-308	83
<b>BR-316</b>	<b>1085</b>

Fonte: Autora

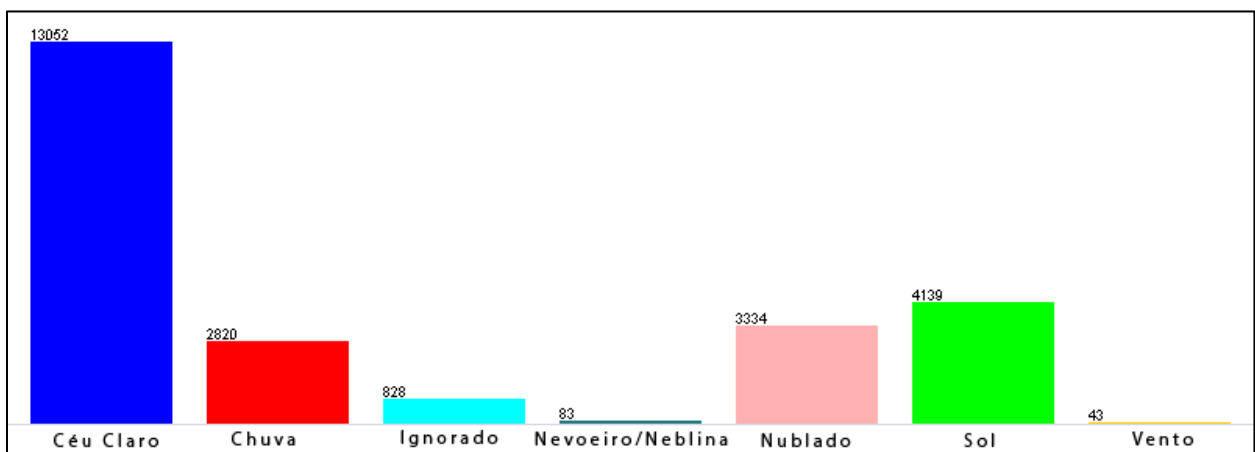
Analisando o traçado da via e a condição meteorológica das ocorrências de trânsito entre 2007 a Setembro de 2018, observou-se que os acidentes de trânsito predominam em linha reta e em céu claro, como mostram as Figuras 24 e Figura 25, respectivamente.

Figura 24 - Traçado da Via nos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018.



Fonte: Autora

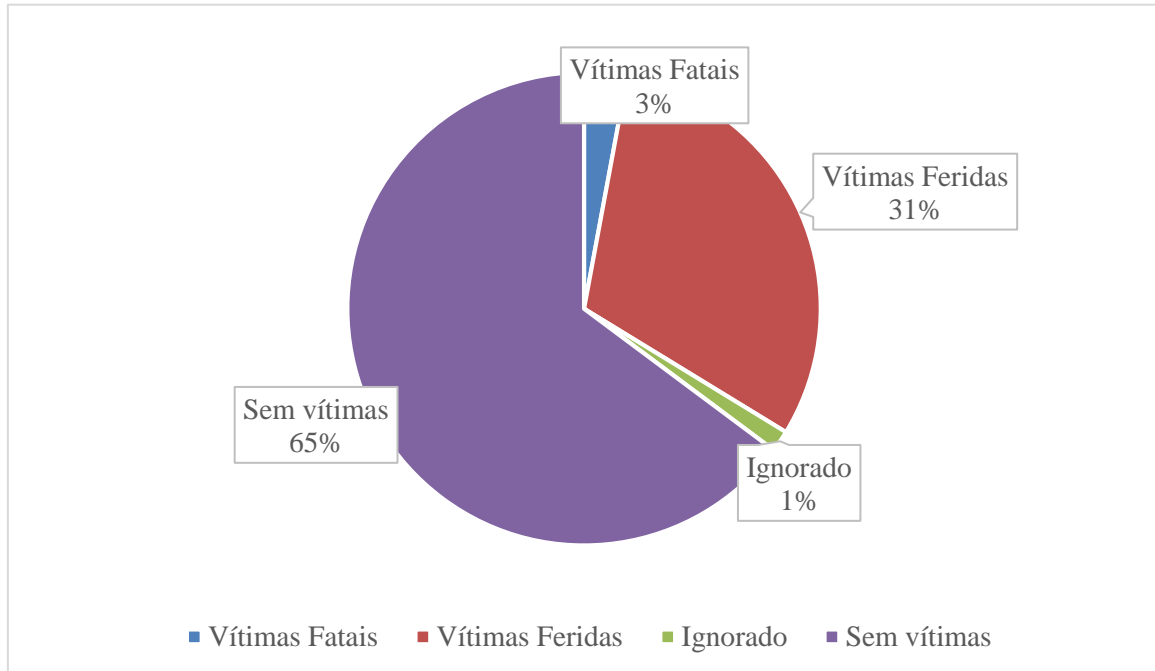
Figura 25 - Condição Meteorológica nos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018



Fonte: Autora

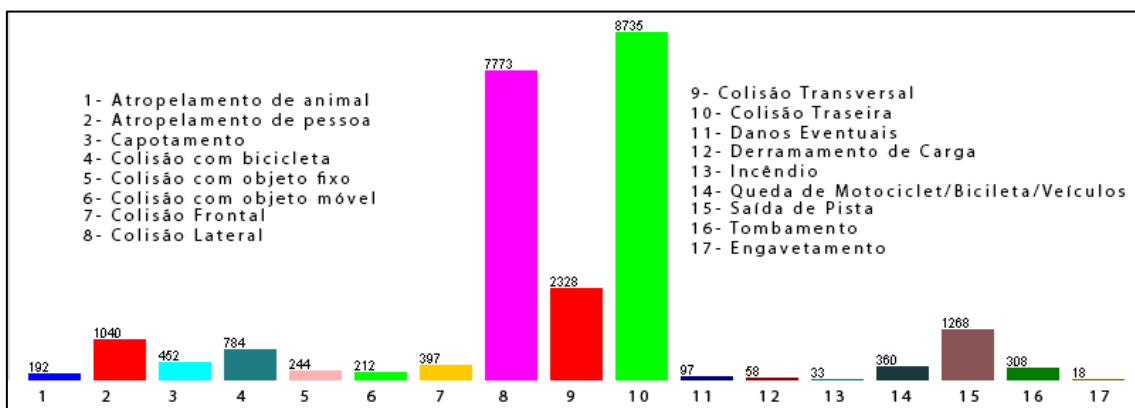
Pela classificação dos acidentes (Figura 26) bem como o tipo de acidente (Figura 27), a estatística mostra que houveram 7.485 vítimas feridas (31%) e que há um predomínio de colisões do tipo traseira e lateral, como mostram as Figuras abaixo.

**Figura 26** - Classificação dos Acidentes de trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018



**Fonte:** Autora

**Figura 27** - Tipo de Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018

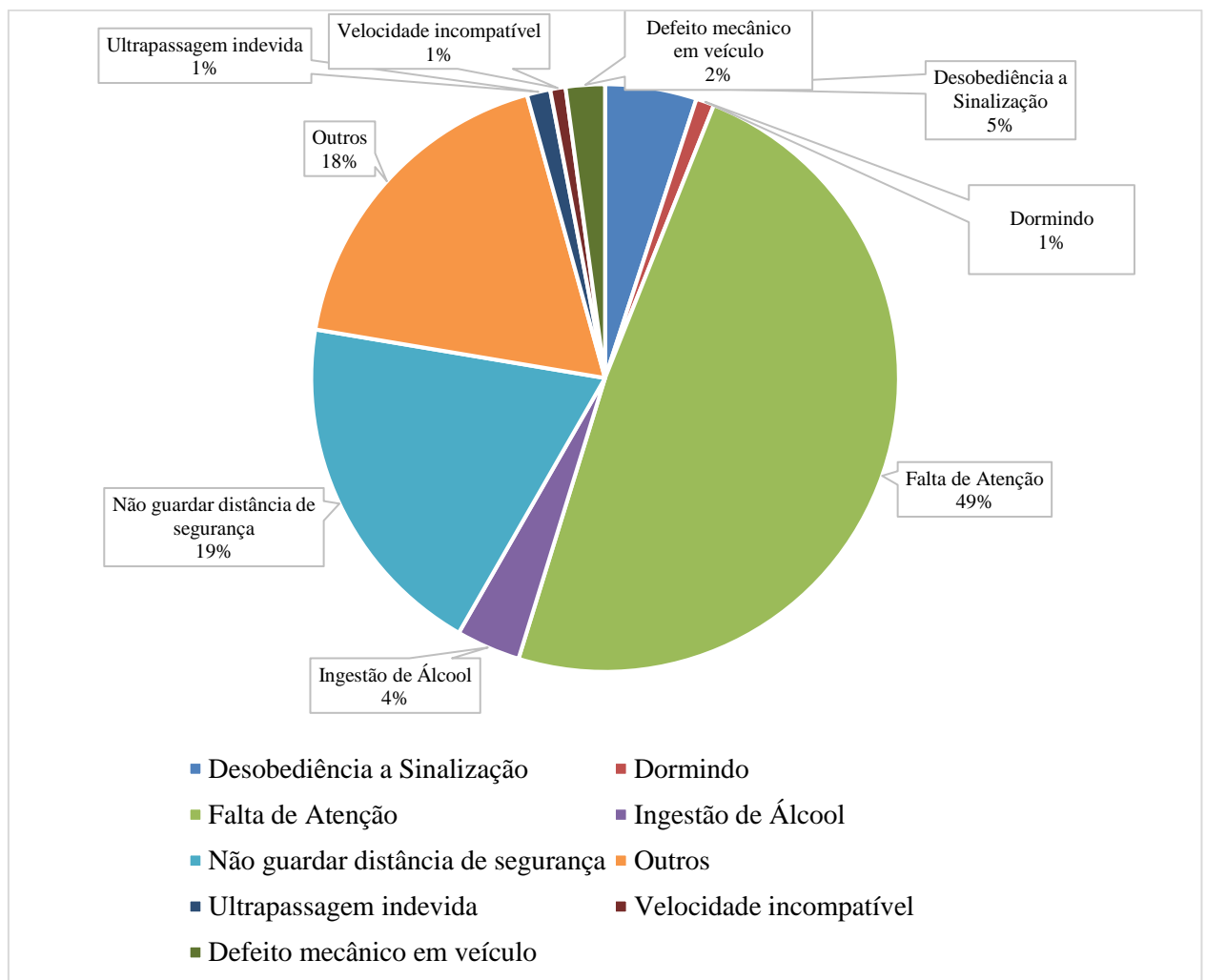


**Fonte:** Autora

A análise estatística mostrou que a grande causa dos acidentes de trânsito (Figura 28) é por falta de atenção dos condutores. Há muitas razões para o aumento dos acidentes de trânsito, como: urbanização rápida, padrões de segurança deficientes, falta de fiscalização, pessoas dirigindo distraídas ou fatigadas, outras sob a influência de drogas ou álcool, excesso de

velocidade e o não uso de cintos de segurança ou capacetes. A alta dependência dos aparelhos eletrônicos nos dias de hoje, bem como a poluição visual são fatores que podem tirar a atenção do condutor no volante e podem, assim, causar ocorrências de acidentes de trânsito.

**Figura 28** - Causa dos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018



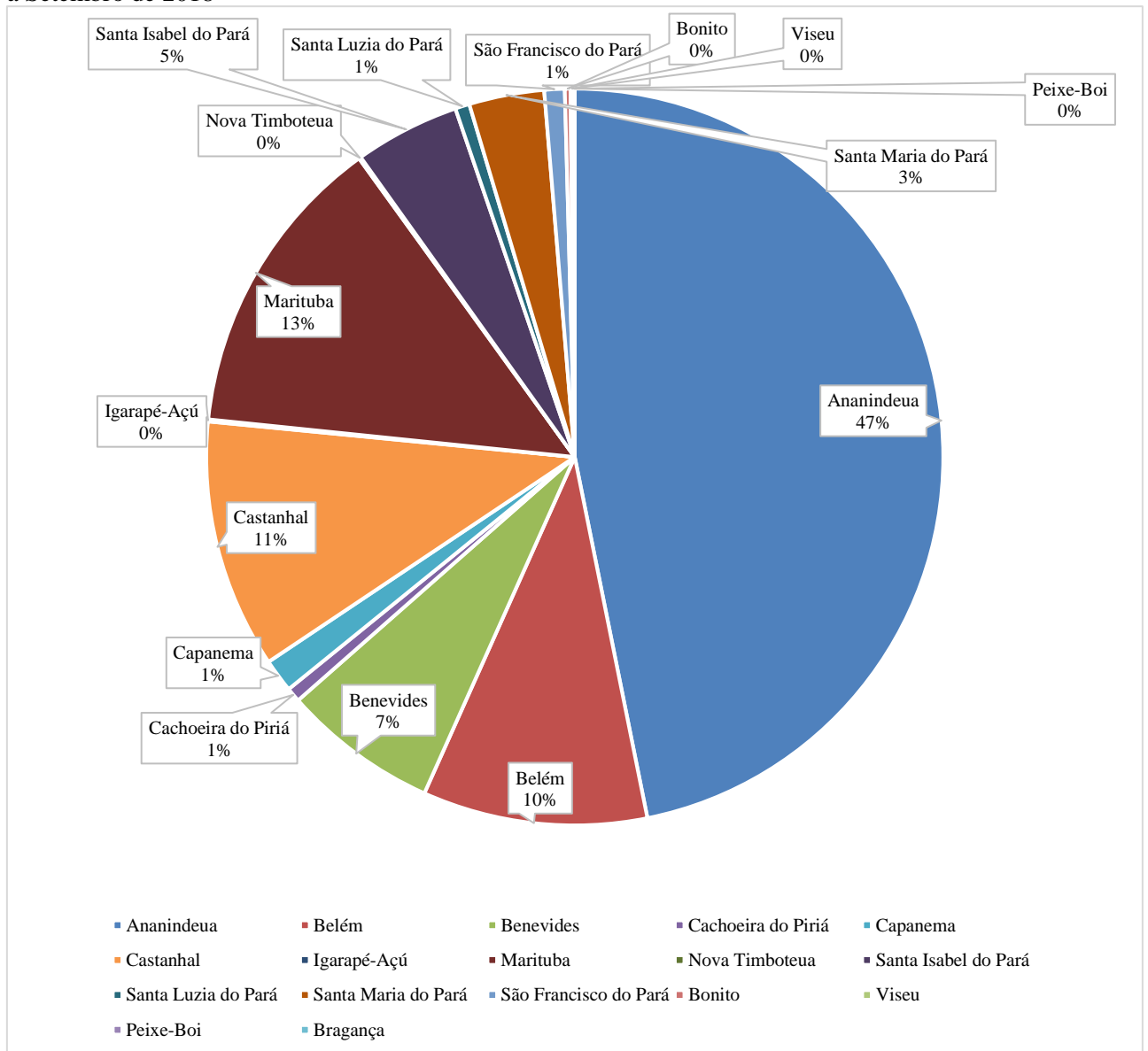
**Fonte:** Autora

Analisando os municípios (Figura 29) com mais índices de acidentes, Ananindeua com 475 mil habitantes e Marituba com 108 mil habitantes lideram no *ranking*. Esse trecho entre os dois municípios é altamente perigoso porque foram erguidos com o decorrer do tempo grandes conglomerados urbanos ao longo das margens da rodovia, além de ser a única saída por terra da capital Belém.

Outro fator que pode explicar o grande número de acidentes é o trânsito intenso no local. A rodovia recebe muitos veículos diariamente de moradores de Ananindeua e Marituba

que trabalham na capital do Pará. Provavelmente, se houvessem rodovias alternativas para essas regiões, o número de acidentes poderia ser menor.

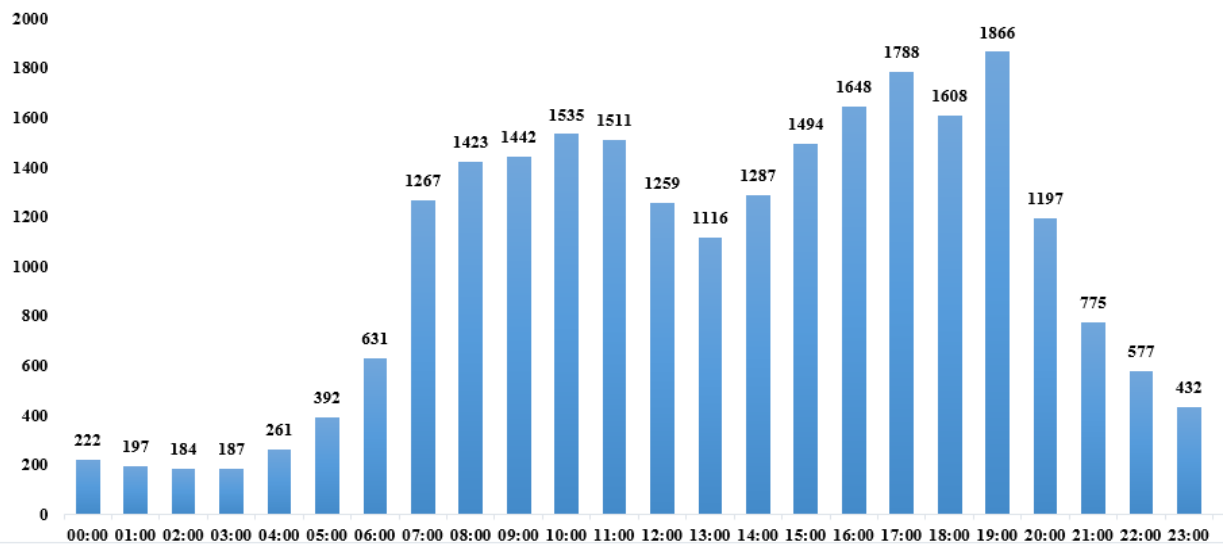
**Figura 29** - Municípios dos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018



**Fonte:** Autora

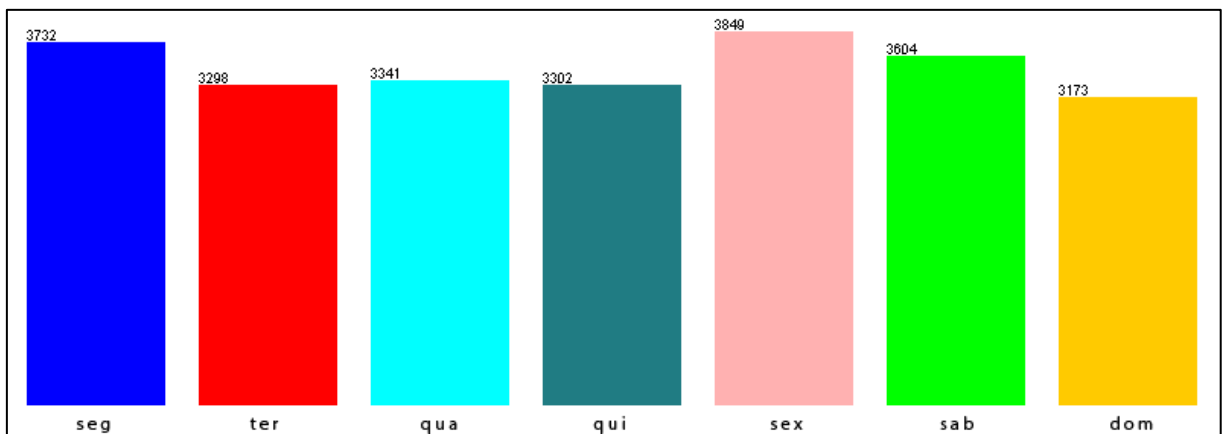
Com relação aos horários e dia da semana com mais ocorrências de acidentes, observou-se que entre 13:00 e 17:00 horas é a faixa de horário com mais probabilidade de acidentes de trânsito (Figura 30), e 19:00 é o horário com mais acidentes de trânsito. Sexta-feira e segunda-feira são os dias da semana com mais ocorrências, como mostra a Figura 31.

**Figura 30** - Horário dos acidentes de trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018



**Fonte:** Autora

**Figura 31** - Dia da semana dos acidentes de trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018

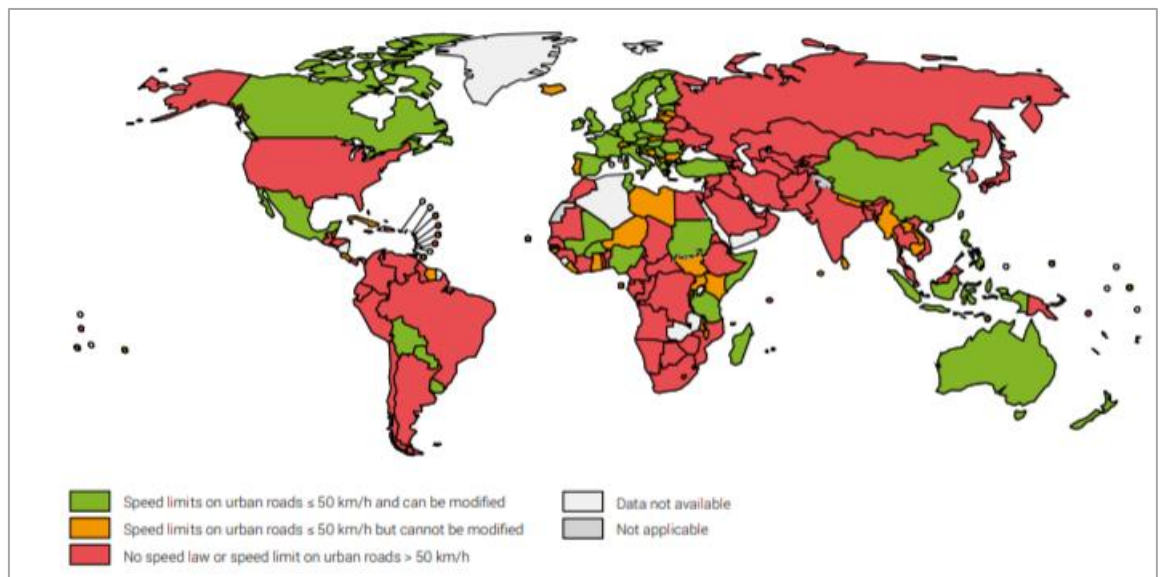


**Fonte:** Autora

Quando verifica-se os quilômetros com mais acidentes, os quilômetros 8 e 0 são os mais perigosos. O alto fluxo de pessoas e veículos podem explicar o alto índice. A Fotografia 33 foi tirada ao longo do quilômetro 8, em Ananindeua. Percebe-se um conglomerado de lojas, condomínios residenciais, vendedores ambulantes e conseqüentemente um alto fluxo de pessoas. De acordo com o Diretor Geral da Organização Mundial da Saúde, Dr Tedros Adhanom Ghebreyesus, os acidentes de trânsito não são “acidentes”, são completamente evitáveis quando há implantação da segurança rodoviária - uma questão que não recebe nada perto da atenção que merece - e é uma grande oportunidade para salvar vidas em todo o mundo.

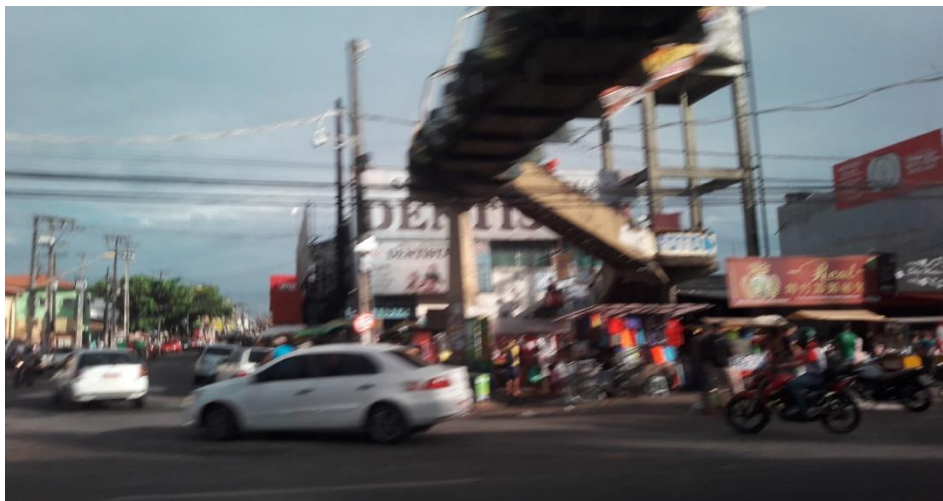
O limite de velocidade nas áreas urbanas no Brasil precisa ser melhorado. De acordo com o último relatório global divulgado pela OMS em Dezembro de 2018 sobre o estado da segurança viária, o Brasil se enquadrava na legenda nenhuma lei de velocidade ou velocidade em via urbana maior que 50 km/h (Figura 32), ou seja, é um fator que precisa ser revisto na legislação brasileira. A Figura 34 é uma distribuição do Weka antes dos valores serem agrupados. Observa-se que havia 880 valores diferentes, variando do quilômetro 0 ao 270.

**Figura 32** - Países com legislação de velocidade que cumprem as melhores práticas, 2017



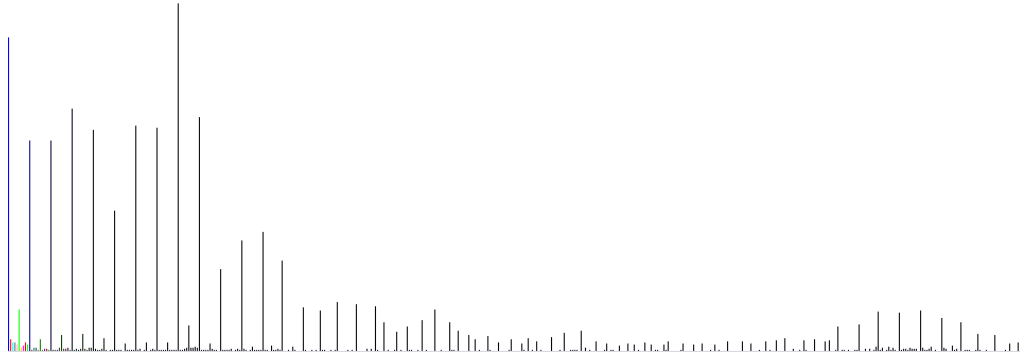
Fonte: OMS, 2018.

**Figura 33** – Fotografia no km 8 da Rodovia BR-316



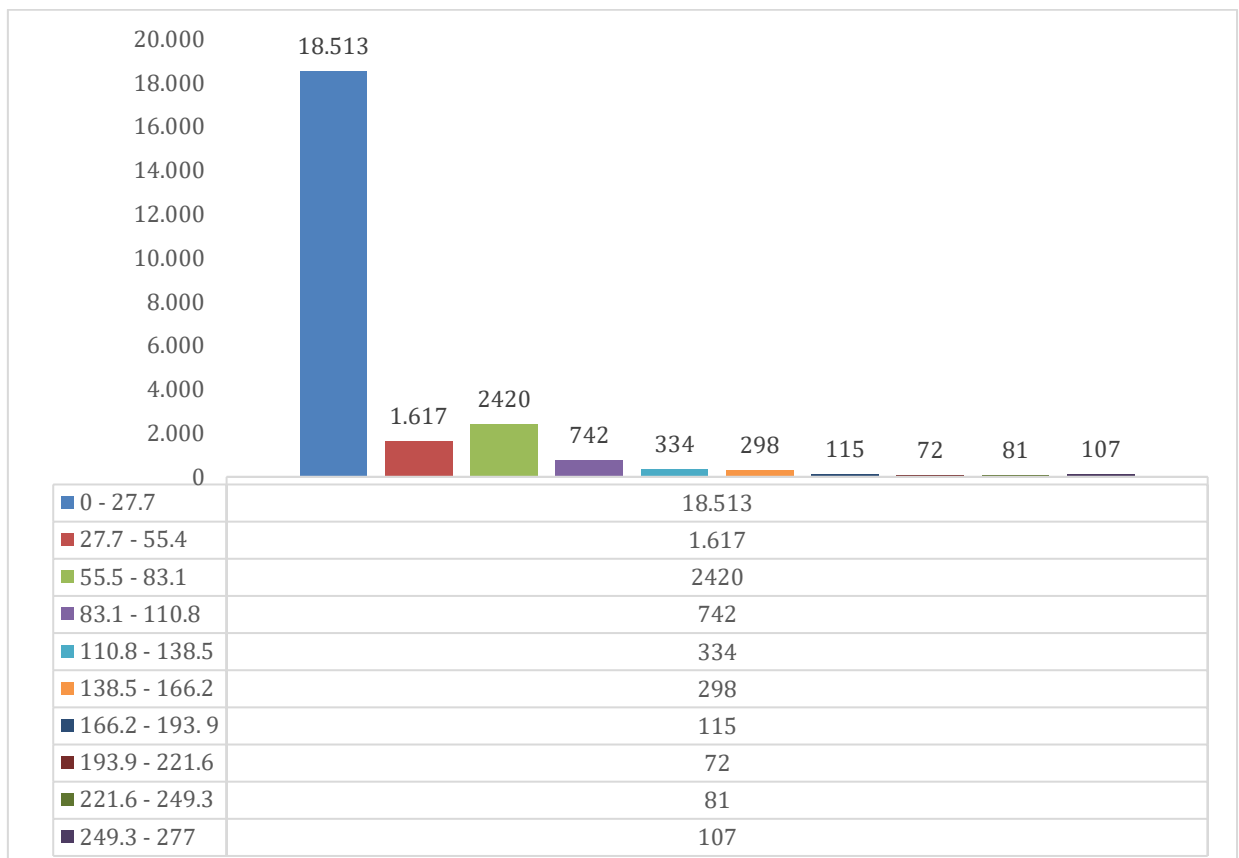
Fonte: Autora

**Figura 34** - Distribuição dos Quilômetros antes da aplicação do filtro Discretize



Depois de ser aplicado o filtro Discretize (explicado na página 38), o Weka fez uma distribuição (Figura 35) com 10 intervalos. Analisando a faixa dos quilômetros entre 0 a 27.7 (Centro de Ananindeua até Benevides), observa-se que é a mais tendenciosa a acidentes de trânsito na BR-316, pois houveram 18.513 registros de ocorrências de trânsito.

**Figura 35** - Distribuição dos quilômetros dos acidentes de trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018



Fonte: Autora

Além destes resultados, o apêndice C apresenta outros gráficos gerados a partir da mineração de dados no Weka.

## 4.2 RESULTADOS COM O ALGORITMO APRIORI ENVOLVENDO OS ATRIBUTOS HORÁRIO, KM, MUNICÍPIO E CAUSA DO ACIDENTE

Foram definidos dois cenários para uma análise e a partir deste foram feitos alguns testes. A Figura 36 é um teste de uma aplicação do algoritmo Apriori. Nesse exemplo há 30 regras de associações. A regra 21, por exemplo, mostra uma tendência de acidentes entre 18:00 e 23:00h no km 8 da BR-316 no município de Ananindeua. Nesta regra houve confiabilidade de 99%. A cobertura de uma regra de associação é o número de instâncias para as quais prediz corretamente - isso é frequentemente chamado de suporte. Sua precisão - geralmente chamada confiança - é o número de instâncias que ele prevê corretamente, expressas como proporção de todas as instâncias às quais se aplica.

**Figura 36** - Teste de algoritmo Apriori no Weka com dados de Acidente de Trânsito da Polícia Rodoviária Federal na BR-316

```

Minimum support: 0.01 (243 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 77

Size of set of large itemsets L(2): 235

Size of set of large itemsets L(3): 65

Best rules found:

1. horario=seis_a_onze_horas km=km 8 733 ==> municipio=ananindeua 733 <conf:(1)> lift:(2.14) lev:(0.02) [389] conv:(389.68)
2. horario=seis_a_onze_horas km=km 7 439 ==> municipio=ananindeua 439 <conf:(1)> lift:(2.14) lev:(0.01) [233] conv:(233.38)
3. horario=meio_dia_a_dezessete_horas km=km 4 398 ==> municipio=ananindeua 398 <conf:(1)> lift:(2.14) lev:(0.01) [211] conv:(211.59)
4. dia_semana=sexta km=km 8 372 ==> municipio=ananindeua 372 <conf:(1)> lift:(2.14) lev:(0.01) [197] conv:(197.77)
5. dia_semana=quinta km=km 8 300 ==> municipio=ananindeua 300 <conf:(1)> lift:(2.14) lev:(0.01) [159] conv:(159.49)
6. dia_semana=terca km=km 8 286 ==> municipio=ananindeua 286 <conf:(1)> lift:(2.14) lev:(0.01) [152] conv:(152.05)
7. km=km 63 273 ==> municipio=castanhal 273 <conf:(1)> lift:(9.15) lev:(0.01) [243] conv:(243.17)
8. horario=meio_dia_a_dezessete_horas km=km 61 - 69 537 ==> municipio=castanhal 536 <conf:(1)> lift:(9.14) lev:(0.02) [477] conv:(239.16)
9. horario=meio_dia_a_dezessete_horas km=km 7 482 ==> municipio=ananindeua 481 <conf:(1)> lift:(2.13) lev:(0.01) [255] conv:(128.12)
10. horario=dezoito_a_vinte_e_tres_horas km=km 61 - 69 348 ==> municipio=castanhal 347 <conf:(1)> lift:(9.13) lev:(0.01) [308] conv:(154.99)
11. km=km 61 - 69 1375 ==> municipio=castanhal 1371 <conf:(1)> lift:(9.13) lev:(0.05) [1220] conv:(244.95)
12. dia_semana=quarta km=km 8 285 ==> municipio=ananindeua 284 <conf:(1)> lift:(2.13) lev:(0.01) [150] conv:(75.76)
13. horario=meio_dia_a_dezessete_horas km=km 8 769 ==> municipio=ananindeua 766 <conf:(1)> lift:(2.13) lev:(0.02) [405] conv:(102.21)
14. horario=seis_a_onze_horas km=km 61 - 69 430 ==> municipio=castanhal 428 <conf:(1)> lift:(9.11) lev:(0.02) [381] conv:(127.67)
15. km=km 8 2018 ==> municipio=ananindeua 2008 <conf:(1)> lift:(2.12) lev:(0.04) [1062] conv:(97.53)
16. km=km 7 1236 ==> municipio=ananindeua 1229 <conf:(0.99)> lift:(2.12) lev:(0.03) [650] conv:(82.14)
17. horario=seis_a_onze_horas km=km 5 326 ==> municipio=ananindeua 324 <conf:(0.99)> lift:(2.12) lev:(0.01) [171] conv:(57.77)
18. horario=seis_a_onze_horas km=km 3 464 ==> municipio=ananindeua 461 <conf:(0.99)> lift:(2.12) lev:(0.01) [243] conv:(61.67)
19. dia_semana=segunda km=km 8 275 ==> municipio=ananindeua 273 <conf:(0.99)> lift:(2.12) lev:(0.01) [144] conv:(48.73)
20. dia_semana=sexta km=km 3 256 ==> municipio=ananindeua 254 <conf:(0.99)> lift:(2.12) lev:(0.01) [134] conv:(45.37)
21. horario=dezoito_a_vinte_e_tres_horas km=km 8 438 ==> municipio=ananindeua 434 <conf:(0.99)> lift:(2.12) lev:(0.01) [228] conv:(46.57)
22. km=km 4 1243 ==> municipio=ananindeua 1231 <conf:(0.99)> lift:(2.11) lev:(0.03) [648] conv:(50.83)
23. horario=dezoito_a_vinte_e_tres_horas km=km 4 306 ==> municipio=ananindeua 303 <conf:(0.99)> lift:(2.11) lev:(0.01) [159] conv:(40.67)
24. km=km 3 1414 ==> municipio=ananindeua 1400 <conf:(0.99)> lift:(2.11) lev:(0.03) [737] conv:(50.11)
25. km=km 5 796 ==> municipio=ananindeua 787 <conf:(0.99)> lift:(2.11) lev:(0.02) [414] conv:(42.32)
26. horario=dezoito_a_vinte_e_tres_horas km=km 3 442 ==> municipio=ananindeua 437 <conf:(0.99)> lift:(2.11) lev:(0.01) [229] conv:(39.16)
27. horario=seis_a_onze_horas km=km 6 504 ==> municipio=ananindeua 498 <conf:(0.99)> lift:(2.11) lev:(0.01) [261] conv:(38.28)
28. horario=meio_dia_a_dezessete_horas km=km 6 417 ==> municipio=ananindeua 412 <conf:(0.99)> lift:(2.11) lev:(0.01) [216] conv:(36.95)
29. dia_semana=sabado km=km 8 314 ==> municipio=ananindeua 310 <conf:(0.99)> lift:(2.11) lev:(0.01) [162] conv:(33.39)
30. km=km 6 1246 ==> municipio=ananindeua 1229 <conf:(0.99)> lift:(2.11) lev:(0.03) [645] conv:(36.8)

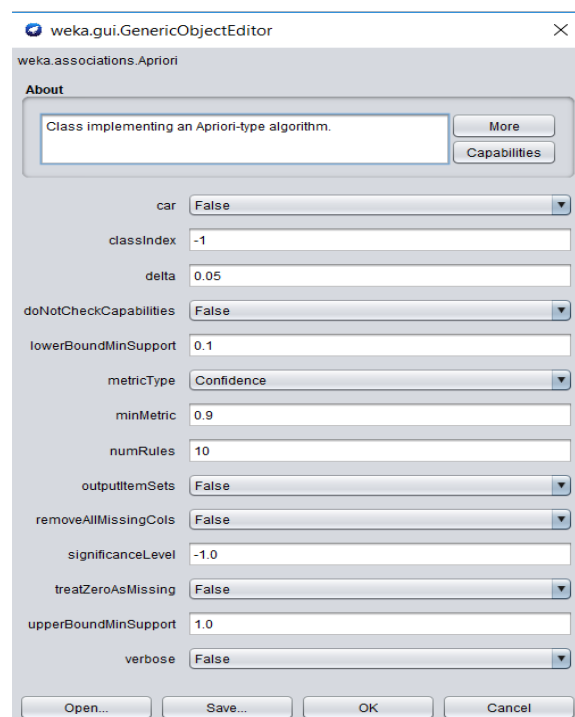
```

Fonte: Autora

Os parâmetros, conforme documentação do Weka dessa interface são descritos abaixo:

- **car**: Se habilitado, as regras de associação de classe são extraídas em vez das regras de associação (gerais).
- **classIndex**: Índice do atributo de classe. Se definido como -1, o último atributo é considerado atributo de classe.
- **delta**: Diminuir iterativamente o suporte por este fator. Reduz o suporte até que o suporte mínimo seja atingido ou o número necessário de regras tenha sido gerado.
- **lowerBoundMinSupport**: Limite inferior para suporte mínimo.
- **metricType**: Define o tipo de métrica pelo qual se pode classificar as regras.
- **minMetric**: Pontuação métrica mínima.
- **numRules**: Número de regras para encontrar.
- **outputItemSets**: Se ativada, os conjuntos de itens também são enviados.
- **removeAllMissingCols**: Remove colunas com todos os valores ausentes.
- **significanceLevel**: Nível de significância. Teste de significância (apenas métrica de confiança).
- **upperBoundMinSupport**: Limite superior para suporte mínimo. Comece a diminuir iterativamente o suporte mínimo deste valor.
- **verbose**: Se habilitado, o algoritmo será executado no modo detalhado.

**Figura 37** – Parâmetro padrão do algoritmo Apriori no Weka



**Fonte:** Autora

Desde do início do projeto, foram trabalhados quase todos os atributos contidos na planilha da Polícia Rodoviária Federal. Sendo assim, foram feitos testes com vários atributos por regras de associação, bem como de classificação. Entretanto, houve um resultado melhor usando o algoritmo Apriori.

Em um primeiro cenário foram usados os atributos horário, km, município e causa do acidente para verificar os padrões com esses atributos. A Figura 38 mostra o cenário 1.

**Figura 38** - Cenário 1 – atributos: horário, km, município e causa do acidente

```

=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    ACIDENTES_DE_TRANSITO-weka.filters.unsupervised.attribute.Discretize-0-B10-M-1.0-R6-precision6-weka.filters.unsupervised.attribute.NumericToNominal
Instances:   24299
Attributes:  4
             horario
             km
             municipio
             causa_acidente
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (2430 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Size of set of large itemsets L(2): 16

Size of set of large itemsets L(3): 8

Best rules found:

1. horario=06:00-11:00 municipio=ananindeua 4107 ==> km=0 - 27.7 4094 <conf:(1)> lift:(1.31) lev:(0.04) [964] conv:(69.85)
2. municipio=ananindeua causa_acidente=falta_de_atencao 6034 ==> km=0 - 27.7 6009 <conf:(1)> lift:(1.31) lev:(0.06) [1411] conv:(55.26)
3. municipio=ananindeua 11381 ==> km=0 - 27.7 11328 <conf:(1)> lift:(1.31) lev:(0.11) [2657] conv:(50.19)
4. horario=18:00-23:00 municipio=ananindeua 2814 ==> km=0 - 27.7 2798 <conf:(0.99)> lift:(1.31) lev:(0.03) [654] conv:(39.42)
5. horario=12:00-17:00 municipio=ananindeua 3886 ==> km=0 - 27.7 3862 <conf:(0.99)> lift:(1.3) lev:(0.04) [901] conv:(37.01)
6. municipio=ananindeua causa_acidente=nao_guardar_distancia_de_seguranca 2525 ==> km=0 - 27.7 2509 <conf:(0.99)> lift:(1.3) lev:(0.02) [585] conv:(35.37)
7. municipio=marituba 3259 ==> km=0 - 27.7 3233 <conf:(0.99)> lift:(1.3) lev:(0.03) [750] conv:(28.74)

```

Para a leitura dos dados, Markov & Larose (2007) afirmam que o conhecimento descoberto é representado na forma de regras [Se... então]. Essas são representadas da seguinte forma “Se os atributos preditivos de uma dupla satisfazem as condições no antecedente da regra, então a dupla tem a classe indicada no conseqüente da regra”. Dessa forma, a regra: **(horario=06:00-11:00 municipio=ananindeua 4107 ==> km=0 - 27.7 4094 Conf: 1 )**, pode ser lida como: das 4107 ocorrências entre o horário 06:00 e 11:00 h, no município de Ananindeua tiveram como consequência uma maior probabilidade de ocorrência de acidentes os quilômetros entre 0 e 27,7. Neste caso, a confiança é de 100%, pois as 4107 ocorrências que possuíam os valores dos atributos analisados do conseqüente também possuíam, como resultado, os quilômetros entre 0 e 27,7.

A Tabela 8 abaixo mostra as sete regras geradas com um suporte mínimo de 0.1, confiança mínima de 0.9 (90%), um mínimo de 10 regras e um total de 18 combinações.

**Tabela 8** – Regras de Associação encontradas no Cenário 1

<b>Número</b>	<b>Resultado</b>	<b>Confiança%</b>
1	horario=06:00-11:00 municipio=ananindeua 4107 ==> km=0 - 27.7 4094	100
2	municipio=ananindeua causa_acidente=falta_de_atencao 6034 ==> km=0 - 27.7 6009	100
3	municipio=ananindeua 11381 ==> km=0 - 27.7 11328	100
4	horario=18:00-23:00 municipio=ananindeua 2814 ==> km=0 - 27.7 2798	99
5	horario=12:00-17:00 municipio=ananindeua 3886 ==> km=0 - 27.7 3862	99
6	municipio=ananindeua causa_acidente=nao_guardar_distancia_de_seguranca 2525 ==> km=0 - 27.7 2509	99
7	municipio=marituba 3259 ==> km=0 - 27.7 3233	99

#### **4.3 RESULTADOS COM O ALGORITMO APRIORI ENVOLVENDO OS ATRIBUTOS DIA DA SEMANA, MUNICÍPIO, CAUSA DO ACIDENTE E TIPO DE ACIDENTE**

Em um segundo cenário (Figura 39) foi aplicado o algoritmo Apriori com os atributos dia da semana, município, causa do acidente e tipo de acidente.

**Figura 39** - Cenário 2 – atributos: dia da semana, município, causa do acidente e tipo de acidente

```

=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    ACIDENTES_DE_TRANSITO-weka.filters.unsupervised.attribute.Discretize-0-B10-M-1.0-R6-precision6-weka.filters.unsupervised.attribute.NumericToNomin
Instances:   24299
Attributes:  4
             dia_semana
             municipio
             causa_acidente
             tipo_acidente
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (2430 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 15
Size of set of large itemsets L(2): 7
Size of set of large itemsets L(3): 1

Best rules found:

1. causa_acidente=nao_guardar_distancia_de_seguranca 4702 ==> tipo_acidente=colisao_traseira 3821 <conf:(0.81)> lift:(2.26) lev:(0.09) [2130] conv:(3.41)
2. municipio=ananindeua tipo_acidente=colisao_lateral 4138 ==> causa_acidente=falta_de_atencao 2879 <conf:(0.7)> lift:(1.43) lev:(0.04) [863] conv:(1.68)
3. tipo_acidente=colisao_lateral 7773 ==> causa_acidente=falta_de_atencao 5276 <conf:(0.68)> lift:(1.39) lev:(0.06) [1489] conv:(1.6)
4. causa_acidente=falta_de_atencao tipo_acidente=colisao_lateral 5276 ==> municipio=ananindeua 2879 <conf:(0.55)> lift:(1.17) lev:(0.02) [407] conv:(1.17)
5. causa_acidente=nao_guardar_distancia_de_seguranca 4702 ==> municipio=ananindeua 2525 <conf:(0.54)> lift:(1.15) lev:(0.01) [322] conv:(1.15)
6. tipo_acidente=colisao_lateral 7773 ==> municipio=ananindeua 4138 <conf:(0.53)> lift:(1.14) lev:(0.02) [497] conv:(1.14)
7. municipio=ananindeua 11381 ==> causa_acidente=falta_de_atencao 6034 <conf:(0.53)> lift:(1.09) lev:(0.02) [489] conv:(1.09)
8. tipo_acidente=colisao_traseira 8735 ==> municipio=ananindeua 4558 <conf:(0.52)> lift:(1.11) lev:(0.02) [466] conv:(1.11)
9. causa_acidente=falta_de_atencao 11838 ==> municipio=ananindeua 6034 <conf:(0.51)> lift:(1.09) lev:(0.02) [489] conv:(1.08)

```

A Tabela 9 abaixo mostra as nove regras geradas com um suporte mínimo de 0.1, confiança mínima de 50%, houveram 18 combinações.

**Tabela 9** - Regras de Associação encontradas no Cenário 2

Número	Resultado	Confiança %
1	causa_acidente=nao_guardar_distancia_de_seguranca 4702 ==> tipo_acidente=colisao_traseira 3821	81
2	municipio=ananindeua tipo_acidente=colisao_lateral 4138 ==> causa_acidente=falta_de_atencao 2879	70
3	tipo_acidente=colisao_lateral 7773 ==> causa_acidente=falta_de_atencao 5276	68
4	causa_acidente=falta_de_atencao tipo_acidente=colisao_lateral 5276 ==> municipio=ananindeua 2879	55
5	causa_acidente=nao_guardar_distancia_de_seguranca 4702 ==> municipio=ananindeua 2525	54

(Continua)

(Continuação)

<b>Número</b>	<b>Resultado</b>	<b>Confiança %</b>
6	tipo_acidente=colisao_lateral 7773 ==> municipio=ananindeua 4138	53
7	municipio=ananindeua 11381 ==> causa_acidente=falta_de_atencao 6034	53
8	tipo_acidente=colisao_traseira 8735 ==> municipio=ananindeua 4558	52
9	causa_acidente=falta_de_atencao 11838 ==> municipio=ananindeua 6034	51

Diante disso, verifica-se que o aprendizado de máquina é uma nova tecnologia para o conhecimento de mineração a partir de dados, na qual está sendo empregada cada vez mais. A partir das regras geradas pelo algoritmo Apriori, conclui-se que a falta de atenção e não guardar distância de segurança são as duas principais causas de acidente no trânsito na BR-316, no Pará. Ananindeua apareceu em muitas regras de associação. É um município que precisa aplicar padrões de segurança rodoviária a fim de reduzir o alto índice de ocorrências de trânsito. É necessário criar cidades e sistemas de transporte que reduzam a alta dependência em carros.

## 5 CONSIDERAÇÕES FINAIS

O trabalho identificou os fatores de contribuição em acidentes rodoviários acontecidos no período de 2007 a Setembro de 2018 da rodovia BR-316, com o uso da descoberta de conhecimento e mineração de dados, por meio de regras de associação geradas pelo software livre Weka e sua implementação do algoritmo Apriori.

O resultado aponta que nos quilômetros iniciais da Rodovia BR-316 no Pará e mais precisamente entre os quilômetros 8 e 0, em pleno dia, nos municípios de Ananindeua, Marituba e Castanhal há um alto índice de acidentes de trânsito.

Na fase do tratamento dos dados foi identificado muitos erros nas planilhas da base de dados online da Polícia Rodoviária Federal. Ficou evidente que deve haver um maior cuidado na coleta de dados, para que os dados primários tenham maior qualidade.

O Weka é um bom software para mineração de dados. A grande dificuldade neste trabalho foi o manuseio da ferramenta em inglês e o pouco tempo de experiência. Foi necessário um tempo significativo para transformar o arquivo CSV para ARFF. Sem dúvidas, essa transformação é o maior obstáculo para quem deseja trabalhar com a ferramenta Weka.

Com o decorrer do trabalho foi importante entender a necessidade de realizar uma boa limpeza nos dados e saber identificar quando há uma necessidade de remover instâncias para a regra não fica tendenciosa.

Ainda segundo os resultados obtidos neste estudo, sugerem-se alguns temas para trabalhos futuros:

- Utilizar outros algoritmos de mineração de dados em busca de padrões diferentes daqueles obtidos por meio do método Apriori para regras de associação;
- Buscar identificar padrões diferentes para outras rodovias federais, comparar e diferenciar padrões entre as BR-010, BR-153 e BR-230, por exemplo.
- Utilizar e contrastar outras fontes de informação, como dados hospitalares de vítimas de acidentes rodoviários.
- Utilizar a técnica de segmentação (*clustering*) nos acidentes da BR-316 e criar um sistema de alerta com mensagem *pop-up* quando o usuário de um aplicativo móvel entrar na área de um *cluster* onde haja mais perigo de acidentes, por exemplo.

## REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, Rakesh et al. **Fast algorithms for mining association rules**. In: INT. CONF. VERY LARGE DATA BASES, VLDB, 20., 1994, Santiago. Proceedings... Santiago: [s. n.], 1994. p. 487-499.

ANDRADE, E. M; RAMOS, E.M.L.S; ALMEIDA, S.S; ARAÚJO, A.R; PINHEIRO, R.F. Abordagem estatística dos acidentes de trânsito fatais ocorridos em rodovia federal do Estado do Pará. **Revista Brasileira de Segurança Pública**. São Paulo Ano 5 Edição 9 Ago/Set 2011.

AREND, M. **50 Anos de industrialização do Brasil (1955-2005): uma análise evolucionária**. 251f. Tese (Doutorado em Economia). Faculdades de Ciências Econômicas. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

BASTOS, J.T. **Geografia da mortalidade no trânsito no Brasil**. 2011. 150p. Dissertação (Mestrado em Engenharia de Transportes) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

BECKER, K.B. **Amazônia: Geopolítica na virada do III milênio**. Rio de Janeiro: Garamond, 2009.

Confederação Nacional do Transporte (CNT). **Pesquisa CNT de Rodovias 2017**. Disponível em:<<http://pesquisarodovias.cnt.org.br/Downloads/Edicoes//2017/Relat%C3%B3rio%20por%20Unidade%20Federativa/PA.pdf>>. Acesso em: 27/11/2018.

CORDIS. **Safety measures focus on road infrastructure**. Germany. 18/06/2010. Disponível em:< [https://cordis.europa.eu/result/rcn/45102\\_en.html](https://cordis.europa.eu/result/rcn/45102_en.html)>. Acesso em 06/12/2018.

CRUZ, A.G; Santos, E.B. **Aplicação da Mineração de Dados para Mitigação dos Acidentes de Trânsito no Município de Goiânia**. 75f. Projeto de Pesquisa (Curso de Bacharelado em Engenharia de Computação). Centro Universitário de Anápolis – Uni EVANGÉLICA, Anápolis, Goiás, 2018.

Departamento de Polícia Rodoviária Federal. **Acidentes**. Disponível em:< <https://www.prf.gov.br/portal/dados-abertos/acidentes>>. Acesso em: 01/12/2018.

DIAS, Maria Madalena. **Um modelo de formalização do processo de desenvolvimento de sistemas de descoberta de conhecimento em banco de dados**. 2001. 197 f. Tese (Doutorado) - Curso de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2001.

ELVIK, R.; HOYE, A.; VAA, T.; SORENSEN, M. **The handbook of road safety measures**. 2th ed. Bingley: Emerald, 2009. 1124p

Flach, P. A., & Lachiche, N. (1999). **Confirmation-guided discovery of first-order rules with Tertius**. Machine Learning, 42, 61–95.

FURTADO, A. **Contexto histórico das rodovias Bernardo Sayão e (BR-010) e Capitão Pedro Teixeira (BR-316)**. Ananindeua, 23/08/2015. Disponível em:< <http://adrielsonfurtado.blogspot.com/2015/08/contexto-historico-da-abertura-das.html>>. Acesso em 25/11/2018.

FURTADO, J. D. (2007). **“Regulação dos transportes rodoviários no desenvolvimento socioeconômico e na integração cultural: as relações de mobilidade do homem no domínio**

e ocupação do seu espaço vital”. Instituto da Biblioteca Nacional e do Livro – Estudos e pesquisas. Praia.

GALVÃO, N. D. **Aplicação da mineração de dados em bancos da segurança e saúde pública em acidentes de transporte**. 2009. 139P. TESE (DOUTORADO EM CIÊNCIAS) - Universidade Federal de São Paulo, SP.

GOVERNO DA AUSTRÁLIA MERIDIONAL. (2011). “**Em direção ao zero juntos: estratégia de segurança no trânsito da Austrália Meridional**”. Disponível em: <[http://www.towardszerotogether.sa.gov.au/\\_\\_data/assets/pdf\\_file/0020/82163/South\\_Australias\\_Road\\_Safety\\_Strategy\\_to\\_2020.pdf](http://www.towardszerotogether.sa.gov.au/__data/assets/pdf_file/0020/82163/South_Australias_Road_Safety_Strategy_to_2020.pdf)>. Acesso em: 06/12/2018.

GUROVITZ, H. **O que cerveja tem a ver com fraldas?** Disponível em :<<https://exame.abril.com.br/revista-exame/o-que-cerveja-tem-a-ver-com-fraldas-m0053931/>>. Acesso em: 12 abr. 2018

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. San Diego, CA: Morgan Kaufmann, 2001. 550 p.

HAUER, E. **Observational before-after studies in road safety**. Bingley: Emerald Group Publishing Limited, 2007. 289p

HOFFMANN, M.H. Comportamento do condutor e fenômenos psicológicos. **Revista Psicologia: Pesquisa e Trânsito**, v.1, n.1, p.17-24. Belo Horizonte, 2005.

IRF - **International Road Federation. Radar - Road Accident Data Recorder Application**. Disponível em: <http://www.irfnet.ch/roadsafety.php?id=104> . Acesso em: 07 nov. 2013

MAHMUD, A.; Zarrinbasha, E. (2010) Intelligent GIS-based Road Accident Analysis and Real-Time Monitoring Automated System using WiMAX/GPRS”, **International Journal of Engineering**, Volume 2: Issue1, pp. 1-7. Disponível em <<http://www.cscjournals.org/csc/manuscript/Journals/IJE/Volume2/Issue1/IJE-7.pdf>>.

MARKOV, Zdravko; LAROSE, Daniel. **Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage**. Hoboken, NJ: John Wiley & Sons, 2007.

MEIRELLES FILHO, J.C.M. **Livro de ouro da Amazônia**. 5 ed. Rio de Janeiro: Ediouro, 2006.

MOUKARZEL, P. E. **A utilização de radares eletrônicos nas rodovias estaduais de Santa Catarina. Especialização em Segurança Pública da Universidade do Sul de Santa Catarina**. Florianópolis, 1999.

NADLER, B., Nadler, F., & Strnad, B. (2012). Road Safety Inspection – **Manual for Conducting RSI**. Vienna: Austrian Road Safety Fund.

NETO, T.O. Rodovia Transamazônica: Falência de um grande projeto geopolítico. **Revista Geonorte**, Edição Especial 3, Manaus, v. 7, n. 1, p. 282-298, 2013.

Organização Mundial da Saúde (OMS), 2013. **Relatório da situação mundial da segurança no trânsito de 2013: apoio à Década de Ação**. Disponível em: [http://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2013/en/](http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/). Acesso em: 05/07/2018.

Organização Mundial da Saúde (OMS), 2015. **Relatório global sobre o estado da segurança viária, 2015**. Disponível

em:<[https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2015/Summary\\_GS\\_RRS2015\\_POR.pdf](https://www.who.int/violence_injury_prevention/road_safety_status/2015/Summary_GS_RRS2015_POR.pdf)>. Acesso em 05/06/2018.

Organização Mundial da Saúde (OMS), 2018. **Relatório global de status sobre a segurança rodoviária de 2018**. Disponível em:

<[https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2018/en/](https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/)>. Acesso em: 20/12/2018.

OLIVEIRA, A.U. A inserção do Brasil no capitalismo monopolista mundial. In: ROSS, J.L.S. (Org.) **Geografia do Brasil**. 6. ed. São Paulo: Editora da Universidade de São Paulo, 2009.

ONUBR. Nações Unidas do Brasil. **Década de Ação pelo Trânsito Seguro 2011-2020 é lançada oficialmente hoje (11) em todo o mundo**. Disponível em: <<http://www.onu.org.br/decada-de-acao-pelo-transito-seguro-2011-2020-e-lancada-oficialmente-hoje-11-em-todo-o-mundo/>>. Acesso em: 20 ago. 2018.

REGO, J.M.; MARQUES, R.M. **Economia Brasileira**. 3º ed. São Paulo: Saraiva, 2006.

REIS, C. V. **O Uso da Descoberta de Conhecimento em Banco de Dados nos Acidentes da BR-381**. Dissertação apresentada ao Curso de Mestrado (Sistemas de Informação e Gestão do Conhecimento). Universidade Fumec. Belo Horizonte – MG, 2014.

RODRIGUES, P.R.A. **Introdução aos sistemas de transporte no Brasil e à logística internacional**. 4. ed. rev. e ampl. São Paulo: Aduaneiras, 2009.

RUD, Olivia Parr. **Data mining cookbook: modeling data for marketing, risk, and customer relationship management**. New York: Willey Computer Publishing, 2001.

SCARLATO, F.C. População e urbanização brasileira. In: ROSS, J.L.S. (Org.) **Geografia do Brasil**. 6. ed. São Paulo: editora da Universidade de São Paulo, 2009.

Scheffer, T. (2001). **Finding association rules that trade support optimally against confidence**. In L. de Raedt, & A. Siebes (Eds.), *Proceedings of the Fifth European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 424–435).

SETRAN. **Rodovia BR-316 será requalificada e transformada em avenida**. Disponível em:< <http://setran.pa.gov.br/site/noticia/179>>. Publicado em 21/05/2018. Acesso em:09/12/2018.

SHAFABAKHSH, G.A; Famili. A; Bahadori. M.S. **GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran**. *Journal of Traffic and Transportation Engineering*. Volume 4, Issue 3, June 2017, Pages 290-299.

SILVA, A.L.D.N. **Prevalência de Fatores Associados à Ocorrência e Severidade de Acidentes com Bicicleta em Porto Alegre**. 2018. 99p. Dissertação (Mestrado em Engenharia de Produção) - Escola de Engenharia da Universidade Federal do Rio Grande do Sul.

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. **Introdução à mineração de dados: com aplicações em R**. 1. ed. Rio de Janeiro: Elsevier, 2016.

Srikant, R., & Agrawal, R. (1996). **Mining sequential patterns: Generalizations and performance improvements**. In P. M. Apers, M. Bouzeghoub, & G. Gardarin (Eds.), *Proceedings of the Fifth International Conference on Extending Database Technology*. Avignon, France. Lecture Notes in Computer Science. Vol. 1057 (pp. 3–17). London: Springer-Verlag.

TAFFAREL, C.M. **Milagre econômico brasileiro (1968-1973)**. 47f. Monografia (Graduação em Ciências Econômicas). Universidade Comunitária da Região de Chapecó, Chapecó, 2010.

The University of Waikato. Attribute-Relation File Format (ARFF). Disponível em:<<https://www.cs.waikato.ac.nz/ml/weka/arff.html>>. Acesso em 05/08/2018.

WEST, D.M; ALLEN, J.R. **How artificial intelligence is transforming the world**. Disponível em: <<https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/>>. Acesso em: 15/12/2018.

WITTEN, I.H; FRANK, E; HALL, M.A. **Data Mining : practical machine learning tools and techniques**. 3. ed. USA: Elsevier, 2011. 665 p.

## APÊNDICES

**Apêndice A - Informações das planilhas baixadas no banco de dados da PRF**

- Data
- Dia da semana
- Horário
- Município
- UF
- BR
- KM
- Causa do acidente
- Tipo de Acidente
- Classificação do Acidente
- Fase do dia
- Sentido da Via
- Condição Meteorológica
- Tipo de Pista
- Traçado da Via
- Uso do Solo
- Tipo do Veículo
- Marca
- Ano de fabricação do veículo
- Tipo de envolvido
- Estado físico
- Idade
- Sexo
- Ilesos
- Feridos Leves
- Feridos Graves
- Mortos
- Latitude
- Longitude
- Regional
- Delegacia
- UOP

## Apêndice B - Estrutura usada no arquivo ARFF

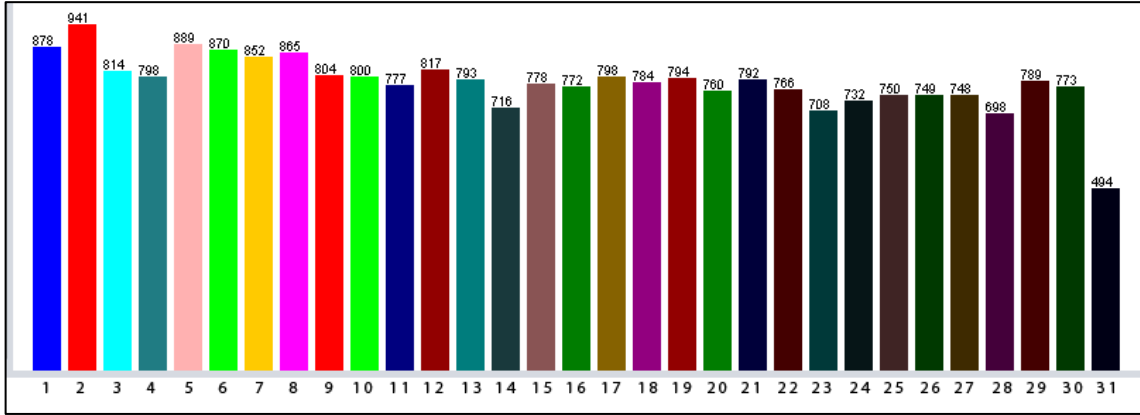
```

@RELATION ACIDENTES_DE_TRANSITO
@ATTRIBUTE dia {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25,
26, 27, 28, 29, 30, 31}
@ATTRIBUTE meses {janeiro, fevereiro, março, abril, maio, junho, julho, agosto, setembro, outubro,
novembro, dezembro}
@ATTRIBUTE ano {2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018}
@ATTRIBUTE dia_semana {segunda, terça, quarta, quinta, sexta, sabado, domingo}
@ATTRIBUTE horario date "HH:mm"
@ATTRIBUTE km numeric
@ATTRIBUTE municipio {ananindeua, belem, benevides, cachoeira_do_piria, capanema, castanhal,
Igarape_açu, marituba, nova_timboteua, santa_isabel_do_para, santa_luzia_do_para,
santa_maria_do_para, sao_francisco_do_para, bonito, viseu, peixe_boi, bragança}
@ATTRIBUTE causa_acidente {animais_na_pista, defeito_na_via, desobediencia_a_sinalização,
dormindo, falta_de_atenção, ingestao_de_alcool, nao_guardar_distancia_de_segurança, outros,
ultrapassagem_indevida, velocidade_incompativel, desobediencia_do_pedreste,
falta_de_atenção_do_pedreste, defeito_mecanico_em_veiculo}
@ATTRIBUTE tipo_acidente {atropelamento_de_animal, atropelamento_de_pessoa, capotamento,
colisao_com_bicicleta, colisao_com_objeto_fixo, colisao_com_objeto_movel, colisao_frontal,
colisao_lateral, colisao_transversal, colisao_traseira, danos_eventuais, derramamento_de_carga,
incendio, queda_de_motocicleta_bicicleta_veiculo, saida_de_pista, tombamento, engavetamento}
@ATTRIBUTE classificacao_acidente {com_vitimas_fatais, com_vitimas_feridas, ignorado,
sem_vitimas}
@ATTRIBUTE fase_dia {amanhecer, anoitecer, pleno_dia, plena_noite}
@ATTRIBUTE sentido_via {crescente, decrescente}
@ATTRIBUTE condicao_meteorologica {ceu_claro, chuva, ignorado, nevoeiro_neblina, nublado, sol,
vento}
@ATTRIBUTE tipo_pista {dupla, multipla, simples}
@ATTRIBUTE tracado_via {cruzamento, curva, reta, desvio_temporario, intersecao_de_vias,
nao_informado, ponte, viaduto, retorno_regulamentado}
@ATTRIBUTE uso_solo {rural, urbano}
@ATTRIBUTE pessoas numeric
@ATTRIBUTE mortos numeric
@ATTRIBUTE feridos_leves numeric
@ATTRIBUTE feridos_graves numeric
@ATTRIBUTE ilesos numeric
@ATTRIBUTE ignorados numeric
@ATTRIBUTE feridos numeric
@ATTRIBUTE veiculos numeric

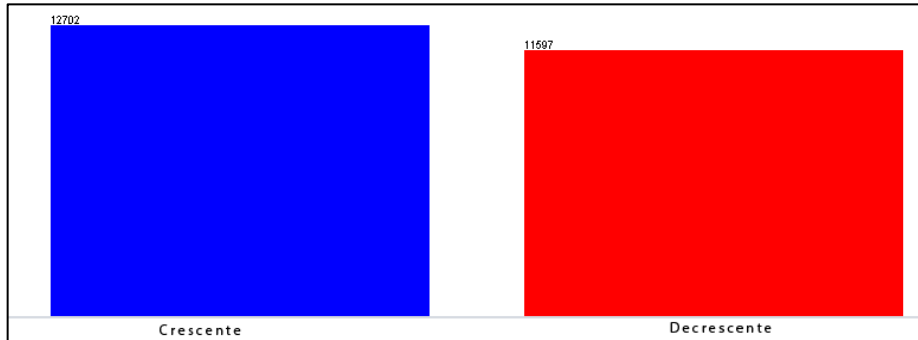
```

Apêndice C – Outros Gráficos gerados no Weka

**Figura 40** - Dia do mês dos Acidentes de Trânsito da BR-316 no estado do PA entre os anos de 2007 a Setembro de 2018



**Figura 41** - Sentido da Via nos Acidentes de Trânsito da BR-316 no estado do PA



**Figura 42** - Fase do dia nos Acidentes de Trânsito da BR-316 no estado do PA

