| **Tecnologia em Análise e Desenvolvimento de Sistemas 5°Período** |  |
| --- | --- |
| **Datamining** |
| **Nomes:**  **Edison Tezolin**  **Eduardo Lucas Veiga de Souza**  **Willian Tezolin Baldessin** |

**DATAMINING**

1. **JUSTIFICAÇÃO DO PRÉ-PROCESSAMENTO E ORIGEM DOS DADOS**

A realização de atividades de Data Mining utilizando fontes de dados relevantes é de suma importância para a obtenção de informações valiosas e insights relevantes. Nesse contexto, o presente trabalho tem como objetivo explorar e analisar o conjunto de dados fornecido pela Polícia Rodoviária Federal, denominado "Documento CSV de Acidentes 2023 (Agrupados por pessoa - Todas as causas e tipos de acidentes)", deste arquivo, foram selecionados apenas a **BR-277** dentro dos limites do município de **Curitiba**. A escolha dessa base de dados se deu em virtude da relevância e abrangência das informações contidas, permitindo um estudo aprofundado dos acidentes rodoviários ocorridos ao longo do ano de 2023. A utilização de técnicas de Data Mining possibilitará a identificação de padrões, tendências e correlações ocultas nos dados, auxiliando na compreensão dos fatores determinantes dos acidentes e, consequentemente, na tomada de decisões embasadas em evidências.

1. **VALIDAÇÃO DAS REGRAS**

A validação de dados desempenha um papel fundamental na garantia da qualidade e confiabilidade das informações utilizadas em qualquer contexto analítico. No presente trabalho, os dados foram submetidos a um processo rigoroso de validação por meio da engenharia reversa. Essa abordagem consiste em analisar detalhadamente a estrutura dos dados, sua formatação e relacionamentos, a fim de verificar a consistência e coerência das informações. Através da aplicação de técnicas de engenharia reversa, foi possível identificar e corrigir possíveis erros, inconsistências ou anomalias nos dados, assegurando assim a integridade dos mesmos.

1. **CARREGAMENTO DOS DADOS**

O carregamento dos dados na ferramenta Weka foi uma escolha devido à sua robustez e eficiência para o processamento e análise de conjuntos de dados. O Weka possui uma ampla seleção de algoritmos de mineração de dados, permitindo a aplicação de técnicas avançadas de aprendizado de máquina e descoberta de padrões. Para realizar o cruzamento de dados da nossa base utilizamos o J48, onde a principal ideia por trás desse algoritmo é criar uma árvore de decisão que divide os dados com base nos atributos mais informativos.

A interface intuitiva e amigável da ferramenta simplifica o pré-processamento dos dados, como limpeza, normalização e transformação, tornando essa etapa mais acessível mesmo para usuários com menos experiência. A Weka também oferece recursos de visualização e avaliação de resultados, facilitando a interpretação dos modelos gerados e o acompanhamento do desempenho dos algoritmos utilizados.

1. **DEMONSTRAÇÃO DO PRODUTO**
2. **Regra 01 e 02:**

* **Remoção dos atributos:** BR - MUNICIPIO - USO SOLO
* **Filtro:** CAUSA\_PRINCIPAL
* **Regra:**

causa\_acidente = Ingestao\_de\_alcool\_pelo\_condutor

| fase\_dia = Plena\_Noite: Nao (3.0)

| fase\_dia = Pleno\_dia: Sim (8.0)

| fase\_dia = Anoitecer: Sim (0.0)

| fase\_dia = Amanhecer: Sim (0.0)

causa\_acidente = Demais\_falhas\_mecanicas\_ou\_eletricas: Nao (0.0)

causa\_acidente = Velocidade\_Incompativel

| fase\_dia = Plena\_Noite: Nao (5.0/1.0)

| fase\_dia = Pleno\_dia: Sim (12.0)

**Resultado:**

- De acordo com a primeira regra identificada, a ingestão de álcool pelo condutor é apontada como a principal causa de acidentes durante o dia.

- Conforme a segunda regra identificada, a principal causa indicada é a velocidade incompatível, responsável por provocar os acidentes.

**b) Regra 03:**

* **Remoção dos atributos:** BR - MUNICIPIO - USO SOLO
* **Filtro:** CAUSA\_PRINCIPAL
* **Regra:**

classificacao\_acidente = Com\_Vitimas\_Feridas

| tracado\_via = Reta

| | sentido\_via = Crescente

| | | condicao\_meteorologica = Ceu\_Claro

| | | | fase\_dia = Plena\_Noite: quarta\_feira (18.0)

| | | | fase\_dia = Pleno\_dia

| | | | | idade = 23: terca\_feira (2.0)

| | | | | idade = 26: quarta\_feira (4.0)

| | | | | idade = 28: sexta\_feira (1.0)

| | | | | idade = 29: quinta\_feira (2.0)

| | | | | idade = 32: terca\_feira (2.0)

| | | | | idade = 34: sexta\_feira (12.0)

| | | | | idade = 35: quarta\_feira (2.0)

| | | | | idade = 36: quarta\_feira (2.0)

| | | | | idade = 37: sexta\_feira (1.0)

| | | | | idade = 44: sexta\_feira (12.0)

| | | | | idade = 48: quinta\_feira (2.0)

| | | | | idade = 58: quarta\_feira (4.0)

**Resultado:**

- Acidentes em trânsito reto, sentido crescente, com condição meteorológica de céu claro e fase do dia em pleno dia, onde a idade dos envolvidos é 34 (sexta-feira) e 44 (sexta-feira).

**c) Regra 04:**

* **Remoção dos atributos:** BR - MUNICIPIO - USO SOLO - SENTIDO VIA
* **Filtro:** TIPO\_ACIDENTE
* **Regra:**

classificacao\_acidente = Com\_Vitimas\_Feridas

| fase\_dia = Plena\_Noite

| | condicao\_meteorologica = Ceu\_Claro: Colisao\_traseira (26.0/1.0)

| | condicao\_meteorologica = Nublado: Atropelamento\_de\_Pedestre (3.0/1.0)

| | condicao\_meteorologica = Chuva: Colisao\_traseira (0.0)

| | condicao\_meteorologica = Garoa\_Chuvisco: Colisao\_traseira (0.0)

**Resultado:**

- Durante a noite, sob céu claro, observa-se que a maioria das vítimas feridas está envolvida em acidentes de colisão traseira.

**d) Regra 05:**

* **Remoção dos atributos:** BR - MUNICIPIO - USO SOLO - SENTIDO VIA
* **Filtro:** IDADE
* **Regra:**

sexo = Masculino

| dia\_semana = sexta\_feira

| | ilesos = 1

| | | tipo\_veiculo = Automovel: 44 (15.0/3.0)

| | | tipo\_veiculo = Motoneta: 44 (0.0)

| | | tipo\_veiculo = Motocicleta: 44 (0.0)

| dia\_semana = sabado

| | ilesos = 0

| | tipo\_veiculo = Automovel

| | | tipo\_acidente = Colisao\_com\_objeto: 28 (2.0)

| | | tipo\_acidente = Atropelamento\_de\_Animal: 50 (0.0)

| | tipo\_veiculo = Motoneta: 30 (0.0)

| | tipo\_veiculo = Motocicleta: 30 (15.0)

**Resultado:**

- É possível observar que, nas ocorrências de acidentes registradas às sextas-feiras, a maioria envolve indivíduos do sexo masculino, os quais saíram ilesos. Esses acidentes são predominantemente relacionados a automóveis e ocorrem em indivíduos com aproximadamente 44 anos de idade. Por outro lado, aos sábados, todos os acidentes envolvendo motocicletas resultam em feridos, sendo esses indivíduos geralmente com cerca de 30 anos de idade.

**e) Regra 06:**

* **Remoção dos atributos:** BR - MUNICIPIO - USO SOLO - IDADE
* **Filtro:** CONDICAO\_METEOROLOCIGA
* **Regra:**

classificacao\_acidente = Com\_Vitimas\_Feridas

| dia\_semana = segunda\_feira

| | sentido\_via = Crescente: Nublado (4.0)

| | sentido\_via = Decrescente: Ceu\_Claro (57.0/5.0)

| dia\_semana = terca\_feira

| | fase\_dia = Plena\_Noite: Ceu\_Claro (3.0)

| | fase\_dia = Pleno\_dia

| | | causa\_acidente = Estacionar\_ou\_parar\_em\_local\_proibido: Nublado (0.0)

| | | causa\_acidente = Iluminacao\_deficiente: Nublado (0.0)

| | | causa\_acidente = Condutor\_Dormindo: Nublado (0.0)

| | | causa\_acidente = Ausencia\_de\_reacao\_do\_condutor: Nublado (6.0)

| | | causa\_acidente = Manobra\_de\_mudanca\_de\_faixa

| | | | causa\_principal = Sim: Nublado (6.0)

| | | | causa\_principal = Nao: Ceu\_Claro (2.0)

**Resultado:**

- A regra indica que acidentes com vítimas feridas apresentam diferentes cenários dependendo do dia da semana e dos atributos selecionados. Em particular, acidentes em segundas-feiras com céu claro e direção decrescente têm alta probabilidade de resultar em vítimas feridas. Já em terças-feiras, acidentes durante o dia com falta de reação do condutor em dias nublados, e acidentes devido a manobras de mudança de faixa com causa principal afirmativa em dias nublados, também apresentam probabilidade de envolver vítimas feridas. Essa regra resumida fornece insights sobre os fatores que influenciam a ocorrência de acidentes com vítimas feridas, auxiliando na compreensão e prevenção desses eventos.

**f) Regra 07 e 08:**

* **Remoção dos atributos:** BR - MUNICIPIO - USO SOLO - IDADE
* **Filtro:** SEXO
* **Regra:**

tipo\_envolvido = Passageiro

| tipo\_veiculo = Automovel: Feminino (52.0/11.0)

| tipo\_veiculo = Motoneta: Feminino (0.0)

| tipo\_veiculo = Motocicleta: Feminino (2.0)

| tipo\_veiculo = Nao\_Informado: Feminino (0.0)

| tipo\_veiculo = Semireboque: Feminino (0.0)

| tipo\_veiculo = Reboque: Feminino (0.0)

| tipo\_veiculo = Caminhao\_trator: Feminino (0.0)

| tipo\_veiculo = Caminhonete: Feminino (0.0)

| tipo\_veiculo = Onibus: Feminino (0.0)

| tipo\_veiculo = Caminhao: Feminino (0.0)

| tipo\_veiculo = Bicicleta: Feminino (0.0)

| tipo\_veiculo = Micro\_onibus: Feminino (0.0)

| tipo\_veiculo = Outros: Feminino (0.0)

| tipo\_veiculo = NA: Feminino (0.0)

tipo\_envolvido = Condutor

| estado\_fisico = Lesoes\_Leves

| | condicao\_meteorologica = Ceu\_Claro: Masculino (55.0/2.0)

| | condicao\_meteorologica = Nublado

| | | dia\_semana = segunda\_feira: Masculino (4.0)

| | | dia\_semana = terca\_feira: Feminino (9.0)

| | | dia\_semana = quarta\_feira: Feminino (0.0)

| estado\_fisico = Nao\_Informado: Nao\_Informado (3.0)

| estado\_fisico = Ileso: Masculino (120.0/5.0)

| estado\_fisico = Lesoes\_Graves: Masculino (10.0/1.0)

| estado\_fisico = Obito: Masculino (6.0)

tipo\_envolvido = Testemunha: Nao\_Informado (1.0)

tipo\_envolvido = Pedestre: Masculino (1.0)

tipo\_envolvido = Nao\_Informado: Nao\_Informado (11.0)

**Resultado:**

- A regra indica que, quando os indivíduos envolvidos são passageiros de automóveis, a maioria é do sexo feminino. Por outro lado, quando os envolvidos são condutores, a predominância masculina é observada em diferentes situações: lesões leves em condições de céu claro, ilesos, lesões graves e óbito. Essa regra revela a relação entre o tipo de envolvido, o tipo de veículo, o estado físico e o sexo predominante em cada cenário de acidente.

**g) Regra 09 e 10:**

* **Remoção dos atributos:** BR - MUNICIPIO - USO SOLO - IDADE
* **Filtro:** CAUSA\_ACIDENTE
* **Regra:**

classificacao\_acidente = Com\_Vitimas\_Feridas

| condicao\_meteorologica = Ceu\_Claro

| | dia\_semana = segunda\_feira

| | | causa\_principal = Sim

| | | | tipo\_pista = Simples: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

| | | | tipo\_pista = Dupla: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (12.0)

| | | | tipo\_pista = Multipla: Condutor\_deixou\_de\_manter\_distancia\_do\_veiculo\_da\_frente (4.0)

| | | causa\_principal = Nao: Condutor\_deixou\_de\_manter\_distancia\_do\_veiculo\_da\_frente (36.0/24.0)

| | dia\_semana = quarta\_feira

| | | tipo\_acidente = Colisao\_com\_objeto: Ausencia\_de\_reacao\_do\_condutor (0.0)

| | | tipo\_acidente = Colisao\_traseira: Ausencia\_de\_reacao\_do\_condutor (18.0)

| | | tipo\_acidente = Saida\_de\_leito\_carrocavel: Ausencia\_de\_reacao\_do\_condutor (0.0)

| | | tipo\_acidente = Capotamento: Ausencia\_de\_reacao\_do\_condutor (0.0)

…

| | | tracado\_via = Viaduto: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

| | | tracado\_via = Simples\_Curva: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

| | dia\_semana = sexta\_feira

| | | causa\_principal = Sim

| | | | fase\_dia = Plena\_Noite: Condutor\_deixou\_de\_manter\_distancia\_do\_veiculo\_da\_frente (2.0)

| | | | fase\_dia = Pleno\_dia: Trafegar\_com\_motocicleta\_ou\_similar\_entre\_as\_faixas (10.0/4.0)

| | | | fase\_dia = Anoitecer: Trafegar\_com\_motocicleta\_ou\_similar\_entre\_as\_faixas (0.0)

| | | | fase\_dia = Amanhecer: Trafegar\_com\_motocicleta\_ou\_similar\_entre\_as\_faixas (0.0)

| | | causa\_principal = Nao

| | | | fase\_dia = Plena\_Noite: Velocidade\_Incompativel (2.0)

| | | | fase\_dia = Pleno\_dia: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (20.0/14.0)

| | | | fase\_dia = Anoitecer: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

| | | | fase\_dia = Amanhecer: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

| | dia\_semana = sabado

| | | causa\_principal = Sim

| | | | tracado\_via = Reta: Velocidade\_Incompativel (12.0)

| | | | tracado\_via = Curva: Velocidade\_Incompativel (0.0)

| | | | tracado\_via = Nao\_Informado: Manobra\_de\_mudanca\_de\_faixa (6.0)

| | | | tracado\_via = Desvio\_Temporario: Velocidade\_Incompativel (0.0)

| | | | tracado\_via = Retorno\_Regulamentado: Velocidade\_Incompativel (0.0)

| | | | tracado\_via = Rotatoria: Velocidade\_Incompativel (0.0)

| | | | tracado\_via = Intersecao\_de\_vias: Velocidade\_Incompativel (0.0)

| | | | tracado\_via = Viaduto: Velocidade\_Incompativel (0.0)

| | | | tracado\_via = Simples\_Curva: Velocidade\_Incompativel (0.0)

| | | causa\_principal = Nao: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (24.0/18.0)

| | dia\_semana = domingo: Velocidade\_Incompativel (1.0)

| condicao\_meteorologica = Nublado

| | causa\_principal = Sim

| | | fase\_dia = Plena\_Noite: Pedestre\_cruzava\_a\_pista\_fora\_da\_faixa (3.0/1.0)

| | | fase\_dia = Pleno\_dia

| | | | dia\_semana = segunda\_feira: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (8.0/4.0)

| | | | dia\_semana = terca\_feira: Manobra\_de\_mudanca\_de\_faixa (6.0)

| | | | dia\_semana = quarta\_feira: Manobra\_de\_mudanca\_de\_faixa (0.0)

| | | | dia\_semana = quinta\_feira: Manobra\_de\_mudanca\_de\_faixa (0.0)

| | | | dia\_semana = sexta\_feira: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (2.0)

| | | | dia\_semana = sabado: Manobra\_de\_mudanca\_de\_faixa (0.0)

| | | | dia\_semana = domingo: Manobra\_de\_mudanca\_de\_faixa (0.0)

| | | fase\_dia = Anoitecer: Manobra\_de\_mudanca\_de\_faixa (0.0)

| | | fase\_dia = Amanhecer: Manobra\_de\_mudanca\_de\_faixa (0.0)

| | causa\_principal = Nao: Ausencia\_de\_reacao\_do\_condutor (12.0/6.0)

| condicao\_meteorologica = Chuva: Conversao\_proibida (3.0)

| condicao\_meteorologica = Garoa\_Chuvisco: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

| condicao\_meteorologica = Sol: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (9.0)

| condicao\_meteorologica = Nevoeiro\_Neblina: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

| condicao\_meteorologica = Vento: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

| condicao\_meteorologica = Ignorado: Reacao\_tardia\_ou\_ineficiente\_do\_condutor (0.0)

classificacao\_acidente = Com\_Vitimas\_Fatais

| causa\_principal = Sim: Transitar\_na\_contramao (9.0)

| causa\_principal = Nao: Iluminacao\_deficiente (9.0)

**Resultado:**

- Essa regra apresenta diferentes cenários relacionados à classificação dos acidentes, considerando a presença de vítimas feridas ou fatais.

No caso de acidentes com vítimas feridas, quando a condição meteorológica é "Céu Claro", observa-se uma série de condições relacionadas a diferentes dias da semana e causas principais. Por exemplo, às segundas-feiras, se a causa principal for "Sim" e o tipo de pista for "Dupla", a causa do acidente é atribuída a "Reação tardia ou ineficiente do condutor". No entanto, se a causa principal for "Não", a causa do acidente é atribuída a "Condutor deixou de manter distância do veículo da frente".

Outras combinações são apresentadas para diferentes dias da semana e condições meteorológicas, como quartas-feiras, sextas-feiras e sábados, com causas principais específicas e tipos de acidente associados.

Já no caso de acidentes com vítimas fatais, a presença ou ausência da causa principal determina a ocorrência de diferentes tipos de acidentes. Por exemplo, se a causa principal for "Sim", o acidente é atribuído a "Transitar na contramão". Se a causa principal for "Não", significa que a "Iluminação deficiente" está presente em todos os acidentes, mesmo não sendo fatais.