

Entrega 1 do Projeto Final

Tema: Sinistros de Trânsito em 2007

Alunos: Rafael Augusto de Souza (2134756) e Luiz Guilherme Geron Manfrim Coelho (2134624)

Repositório Git: <https://github.com/LuizGulhermeGeron/Projeto-Final-CD>

1. Dataset:

O dataset escolhido foi Sinistros de Trânsito Agrupados Por Ocorrência em 2007 (<https://dados.gov.br/dados/conjuntos-dados/sinistros-de-transito-agrupados-por-ocorrencia>). Ele contém 127675 entradas e 26 colunas, com informações sobre localização, horário, gravidade e tipo de acidente.

Esse dataset foi complementado com a estimativa da população no ano de 2006 (<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=17283&t=downloads>).

2. Perguntas:

1. Quais fatores (como horário, tipo de pista, condição meteorológica e uso do solo) estão mais associados à ocorrência de acidentes graves nas rodovias federais em 2007?
2. É possível prever a gravidade de um acidente de trânsito com base nas características da ocorrência (condições climáticas, tipo de via, horário e localização)?

3. Justificativa:

Os acidentes de trânsito representam uma das principais causas de morte e lesões no Brasil, com impactos sociais, econômicos e de saúde pública significativos.

Analisar dados de sinistros permite compreender os fatores associados à ocorrência e gravidade dos acidentes, fornecendo evidências para políticas de prevenção e melhoria da segurança viária.

O dataset de 2007, ao agrupar ocorrências em todo o território nacional, oferece uma visão abrangente do comportamento do trânsito nas rodovias federais, possibilitando estudar padrões temporais, geográficos e ambientais que contribuem para os acidentes.

A análise tem potencial de aplicação direta em políticas públicas e planejamento rodoviário. Entre os principais impactos práticos estão: identificação de fatores de risco (quais condições aumentam a probabilidade de acidentes graves), prevenção direcionada (permitir que órgãos responsáveis priorizem recursos em locais, horários e situações de maior risco), planejamento de infraestrutura (apoiar decisões de investimento em sinalização, iluminação, duplicações e manutenção de vias) e criação de modelos preditivos de risco (possibilitar o desenvolvimento de sistemas inteligentes de alerta e monitoramento, que antecipem situações de alto risco com base em condições observadas).

Os maiores beneficiados com essa análise seriam os usuários de vias rodoviárias, que teriam maior noção do risco associado a cada pista e sua condição. Os gestores públicos também poderiam tirar vantagem da análise para a melhor elaboração de medidas públicas e projetos de infraestrutura, reduzindo as condições de risco para esses usuários.

4. Hipóteses testáveis:

1. H1: Fatores ambientais e temporais influenciam a gravidade do sinistro
2. H2: O número de veículos envolvidos e o tipo de pista são preditores significativos da gravidade
3. H3: Modelos de ensemble apresentam desempenho superior na predição da gravidade do acidente