

Caio Eloi Campos

Análise Exploratória e Preditiva de Acidentes de Trânsito em Belo Horizonte com Dados Abertos da BHTrans

Monografia II apresentada ao Instituto de Ciências Exatas e ao Departamento de Ciência da Computação da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador: João Guilherme Maia de Menezes

Departamento de Ciência da Computação
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais

Belo Horizonte, 2026

Resumo

Este trabalho apresenta a segunda etapa da pesquisa sobre acidentes de trânsito em Belo Horizonte, na qual a análise exploratória conduzida na primeira fase é estendida e materializada em um sistema web interativo para visualização e análise preditiva dos dados abertos disponibilizados pela BHTrans. A solução foi concebida sobre uma arquitetura desacoplada em três camadas: um backend em Kotlin com Spring Boot, responsável por expor uma API REST com filtros e agregações; um frontend em React com TypeScript, que apresenta mapas, gráficos e filtros dinâmicos; e um serviço dedicado de aprendizado de máquina em Python com FastAPI. Os dados, referentes ao período de 2016 a 2023, foram consolidados, migrados para um banco SQLite e georreferenciados a partir da conversão de coordenadas UTM para latitude e longitude. Sobre essa base, o sistema disponibiliza um painel que integra a evolução temporal, os horários e dias mais críticos, os tipos de sinistro mais frequentes, os bairros de maior incidência e um mapa interativo. Foram ainda incorporados modelos de previsão de séries temporais e de análise de severidade, permitindo estimar tendências futuras e os fatores associados à fatalidade dos acidentes. Os resultados demonstram o potencial de transformar uma base bruta de dados públicos em informação acionável para gestores, pesquisadores e cidadãos.

Palavras-chaves: análise de dados, segurança viária, acidentes de trânsito, BHTrans, mobilidade urbana, visualização interativa, aprendizado de máquina.

Abstract

This work presents the second stage of the research on traffic accidents in Belo Horizonte, in which the exploratory analysis carried out in the first phase is extended and materialized into an interactive web system for visualization and predictive analysis of the open data provided by BHTrans. The solution was designed on a decoupled three-layer architecture: a backend in Kotlin with Spring Boot, responsible for exposing a REST API with filters and aggregations; a frontend in React with TypeScript, which presents maps, charts and dynamic filters; and a dedicated machine learning service in Python with FastAPI. The data, covering the 2016 to 2023 period, were consolidated, migrated to a SQLite database and georeferenced by converting UTM coordinates to latitude and longitude. On top of this base, the system provides a dashboard that integrates the temporal evolution, the most critical hours and days, the most frequent accident types, the neighborhoods with the highest incidence and an interactive map. Time-series forecasting and severity analysis models were also incorporated, allowing the estimation of future trends and the factors associated with accident fatality. The results demonstrate the potential of transforming a raw base of public data into actionable information for managers, researchers and citizens.

Keywords: data analysis, road safety, traffic accidents, open data, urban mobility, interactive visualization, machine learning.

Sumário

Resumo	2
Abstract	3
Sumário	4
1. Introdução	5
1.1 Da análise exploratória ao sistema web	5
1.2 Visualização e decisão baseada em dados	5
2. Objetivos	6
2.1 Objetivo geral	6
2.2 Objetivos específicos	6
3. Metodologia e desenvolvimento	6
3.1 Visão geral da arquitetura	6
3.2 Camada de dados	7
3.3 Backend: API REST em Kotlin e Spring Boot	8
3.4 Frontend: dashboard em React e TypeScript	8
3.5 Serviço de aprendizado de máquina	10
3.6 Testes e integração contínua	10
4. Resultados	11
5. Discussão	13
6. Conclusão e perspectivas	13
7. Referências	14

1. Introdução

1.1 Da análise exploratória ao sistema web

O trânsito urbano em grandes cidades como Belo Horizonte enfrenta desafios relacionados ao crescimento populacional, à expansão desordenada da malha urbana e ao aumento da frota de veículos, fatores que resultam em congestionamentos e acidentes com impactos diretos na mobilidade, na saúde pública e na qualidade de vida da população. A BHTrans, autarquia municipal responsável pela gestão da mobilidade urbana, disponibiliza publicamente dados abertos sobre acidentes de trânsito, possibilitando análises aprofundadas por parte de pesquisadores, gestores públicos e cidadãos.

A primeira etapa desta pesquisa concentrou-se na coleta, no tratamento e na análise exploratória desses dados, consolidando um conjunto limpo e estruturado e evidenciando padrões preliminares, como a concentração de acidentes em horários de pico e em dias úteis da semana. Esta fase fundamenta e justifica o desenvolvimento de uma ferramenta interativa de visualização, previsto como continuidade do trabalho. Já na segunda etapa continuei esse percurso, transformando os resultados exploratórios em um produto de software concreto: um sistema web que permite explorar os dados de forma intuitiva e que incorpora, adicionalmente, técnicas de análise preditiva.

1.2 Visualização e decisão baseada em dados

A análise e a visualização de dados são ferramentas fundamentais para transformar grandes volumes de informações em conhecimento acessível e acionável. A importância do uso de dados para fundamentar decisões estratégicas e aprimorar políticas públicas é amplamente discutida por Provost e Fawcett (2013), que destacam o papel de técnicas analíticas na transformação de dados brutos em conhecimento acionável. Aplicações interativas e dashboards permitem que essas informações sejam exploradas de maneira intuitiva, facilitando a identificação de pontos críticos e tendências que poderiam passar despercebidas em análises tradicionais, e tornam a comunicação de resultados mais dinâmica, favorecendo tanto o uso técnico por especialistas quanto a compreensão pelo público em geral.

Neste trabalho, propõe-se a construção de um sistema web interativo de visualização e análise dos acidentes de trânsito em Belo Horizonte, partindo dos dados abertos da BHTrans. Além de consolidar e apresentar as visualizações exploratórias, o sistema integra um mapa georreferenciado e modelos preditivos, ampliando o alcance dos resultados e contribuindo

para democratizar o acesso à informação de segurança viária, favorecendo o uso público, institucional e acadêmico dessas informações.

2. Objetivos

2.1 Objetivo geral

Desenvolver um sistema web interativo de visualização e análise dos acidentes de trânsito a partir dos dados públicos disponibilizados pela BHTrans, facilitando a identificação e a compreensão de fenômenos relacionados à segurança viária.

2.2 Objetivos específicos

- Coletar, tratar e organizar os dados disponibilizados pela BHTrans, de forma a garantir sua consistência e usabilidade.
- Apresentar resultados de análises estatísticas básicas a fim de identificar padrões ou resultados como os perfis dos veículos e das pessoas envolvidas nos acidentes, ranquear os logradouros com maior ocorrência de sinistros, entre outros.
- Desenvolver uma ferramenta de visualização interativa dos dados que possa ser disponibilizada para qualquer pessoa a fim de analisar e tomar decisões embasadas sobre estes dados.

3. Metodologia e desenvolvimento

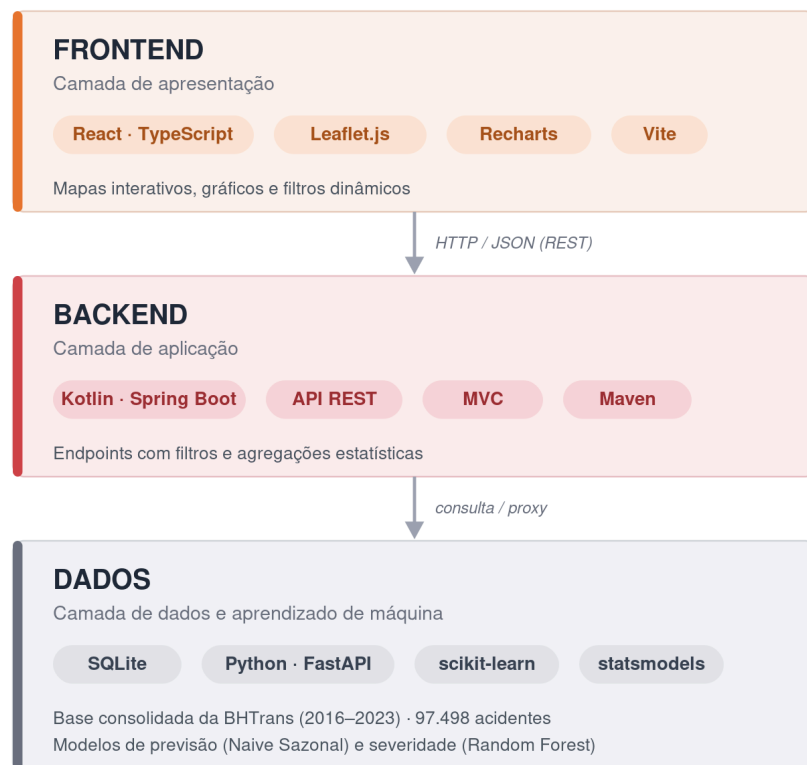
3.1 Visão geral da arquitetura

O desenvolvimento do sistema foi norteado por uma decisão arquitetural fundamental de estruturar a aplicação de forma desacoplada em camadas independentes, em vez de adotar uma solução monolítica baseada em ferramentas de prototipagem como Streamlit ou Dash. Essa decisão, registrada formalmente em um Architecture Decision Record (ADR), considerou critérios como escalabilidade, manutenibilidade, desempenho no processamento dos dados, flexibilidade para evolução futura e demonstração de competências técnicas em um contexto acadêmico. A separação clara de responsabilidades e a busca por um código legível e manutenível alinham-se às boas práticas de engenharia de software descritas por Martin (2008).

A arquitetura adotada organiza-se em três camadas independentes, ilustradas na figura 1. A camada de apresentação é composta pelo frontend em React e TypeScript, com mapas

em Leaflet.js e gráficos em Recharts. A camada de aplicação expõe uma API REST seguindo o padrão MVC, cujos princípios arquiteturais para sistemas distribuídos baseados em rede foram estabelecidos por Fielding (2000).. A camada de dados é sustentada por um banco SQLite, alimentado a partir do dataset consolidado, e por um serviço auxiliar de aprendizado de máquina em Python com FastAPI, scikit-learn e statsmodels.

Figura 1 - Arquitetura do sistema em três camadas independentes



Fonte: Elaborado pelo autor (2026).

3.2 Camada de dados

A base analítica utilizada nesta etapa origina-se do dataset consolidado produzido na primeira fase do projeto, que integrou as diferentes tabelas fornecidas pela BHTrans referentes a acidentes, localizações, veículos e pessoas, conforme resumido na tabela 1. A partir desse conjunto, um script de migração em Python percorre o arquivo CSV consolidado, agrega os registros para uma linha por acidente (identificado pelo número do boletim), extrai campos temporais (ano, mês, hora e dia da semana) e converte as coordenadas originais em UTM (SIRGAS 2000, Zona 23S) para latitude e longitude no sistema WGS84, viabilizando o georreferenciamento. O resultado é gravado em um banco de dados SQLite, consumido diretamente pelo backend.

Tabela 1 - Quantidade de registros por tabela da BHTrans

Tabela	Quantidade
Acidentes	97.498
Localizações	126.879
Veículos	177.705
Pessoas	210.977

Fonte: Elaborado pelo autor (2026).

3.3 Backend - API REST em Kotlin e Spring Boot

O backend foi implementado em Kotlin sobre o framework Spring Boot, organizando-se segundo o padrão MVC com camadas de modelo, repositório, serviço e controlador. A entidade central representa um acidente e reúne atributos temporais, espaciais e descritivos, como tipo de sinistro, regional, bairro, tipo e nome de logradouro, condições de tempo e pavimento, sinalização, velocidade permitida, indicadores de fatalidade e totais de pessoas, vítimas e veículos envolvidos, além das coordenadas geográficas.

A API expõe endpoints REST que permitem listar acidentes com filtros por regional, bairro, tipo, ano e mês; recuperar um acidente específico; obter os pontos georreferenciados para o mapa; e, principalmente, calcular estatísticas agregadas. O endpoint de estatísticas consolida, em uma única resposta, os totais de acidentes e de vítimas fatais e as distribuições por regional, por bairro (os quinze de maior incidência), por tipo, por ano, por hora e por dia da semana, respeitando os filtros aplicados pelo usuário. Essas agregações alimentam diretamente os componentes visuais do frontend.

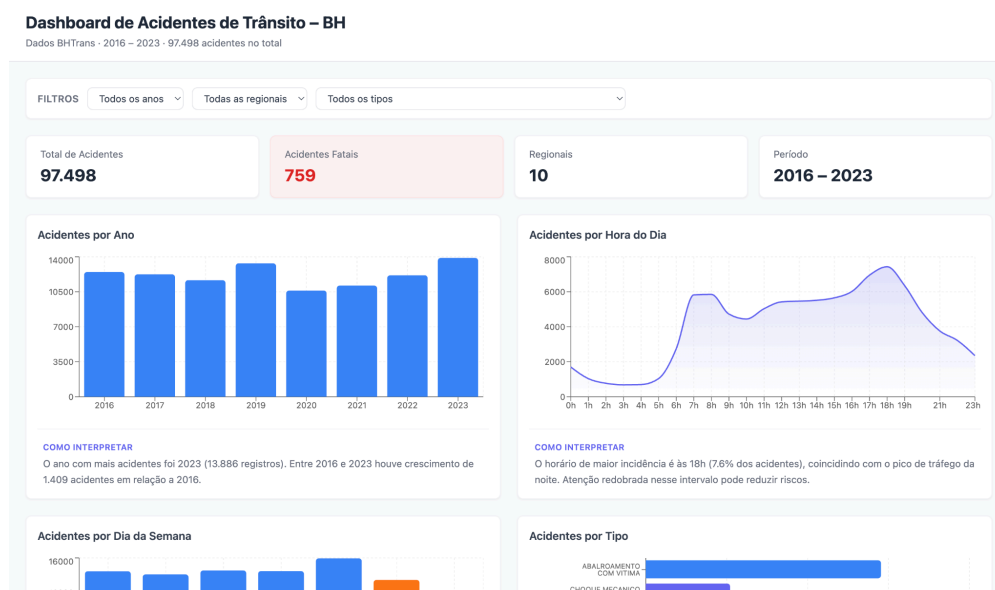
3.4 Frontend - dashboard em React e TypeScript

A interface foi construída em React com TypeScript e empacotada com Vite, consumindo a API por meio de um cliente de serviços. O dashboard apresenta cartões de indicadores e um conjunto de gráficos com evolução anual mostrado na figura 2, distribuição por hora, por tipo, por dia da semana e ranking de bairros, além de um mapa interativo visto na figura 3. Um painel de filtros dinâmicos permite restringir as visualizações por ano, regional e tipo de acidente. Ao alterar um filtro, tanto as estatísticas quanto os modelos

preditivos são recalculados, mantendo todas as visões sincronizadas. Componentes de apoio destacam ainda interpretações automáticas dos gráficos, facilitando a leitura dos padrões.

O mapa interativo foi construído com a biblioteca Leaflet.js sobre tiles do OpenStreetMap, plotando a localização geográfica de cada acidente registrado. A visualização revela a forte concentração espacial na região central de Belo Horizonte, com dispersão ao longo dos principais eixos viários e das conexões com municípios vizinhos. O usuário pode aplicar zoom, navegar pelo mapa e interagir com cada ponto, identificando a regional, o tipo de acidente e se houve vítima fatal.

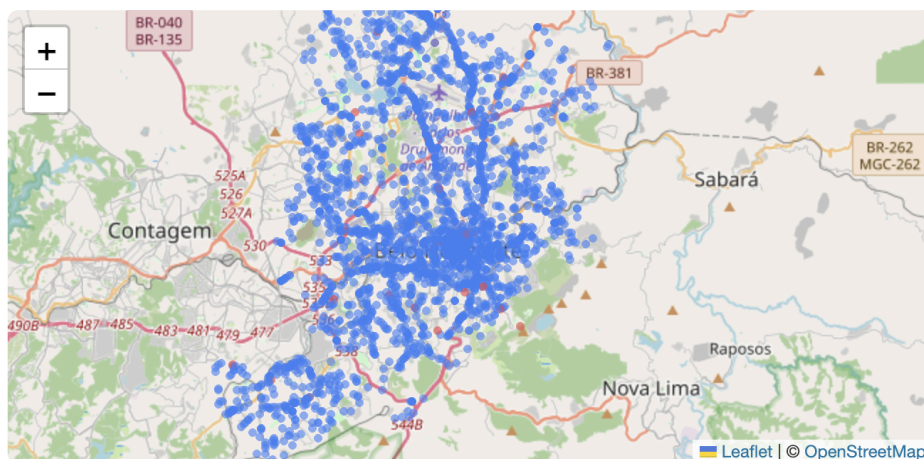
Figura 2 - Visão geral do dashboard interativo



Fonte: Elaborado pelo autor (2026).

Figura 3 - Mapa interativo de acidentes (Leaflet.js + OpenStreetMap)

Localização dos Acidentes



Fonte: Elaborado pelo autor (2026).

3.5 Serviço de aprendizado de máquina

Para incorporar a dimensão preditiva prevista desde a primeira etapa, foi desenvolvido um serviço dedicado de aprendizado de máquina em Python, exposto por meio do framework FastAPI e consumido pelo backend por intermédio de um proxy. Esse serviço lê a mesma base SQLite e disponibiliza endpoints para previsão de séries temporais, análise de severidade e geração do mapa de calor, figura 4, todos sensíveis aos filtros de ano, regional e tipo aplicados na interface.

O modelo de previsão adota a abordagem Naive Sazonal, que assume que cada mês tende a repetir o comportamento observado no mesmo mês dos anos anteriores, capturando a sazonalidade dos acidentes. A projeção dos meses seguintes parte da média histórica de cada mês, e uma banda de incerteza correspondente a mais ou menos um desvio padrão indica o quanto os valores costumam variar. Apesar de simples, o método serve como uma linha de base robusta para comparação com modelos mais complexos.

A análise de severidade emprega um modelo Random Forest, algoritmo que combina centenas de árvores de decisão para estimar o risco de fatalidade de um acidente a partir de variáveis como hora do dia, dia da semana, ano, velocidade permitida e sinalização da via. Além de produzir uma estimativa de risco mais estável do que uma árvore isolada, o modelo quantifica a importância relativa de cada fator, oferecendo subsídios interpretáveis sobre os determinantes da gravidade dos sinistros.

Figura 4 - Mapa de calor - concentração de acidentes por hora e dia da semana



Fonte: Elaborado pelo autor (2026).

3.6 Testes e integração contínua

Para assegurar a qualidade e a confiabilidade do sistema, foram implementados testes automatizados em ambas as camadas principais. Testes de unidade e de integração no backend e testes de componentes no frontend, exercitando os serviços de estatísticas, os controladores e os elementos de interface. O repositório conta ainda com um fluxo de

integração contínua que executa automaticamente a verificação a cada alteração, e o histórico de decisões técnicas é documentado por meio de ADRs, reforçando a rastreabilidade e a reprodutibilidade do desenvolvimento.

4. Resultados

O sistema entrega uma visão consolidada de 97.498 acidentes registrados pela BHTrans entre 2016 e 2023, distribuídos pelas dez regionais de Belo Horizonte, dos quais 759 foram fatais. A partir desses números, o dashboard permite navegar de uma visão panorâmica até o detalhe, explorando a evolução anual, os horários e dias mais críticos, os tipos de sinistro mais frequentes e os bairros de maior incidência, transformando uma base bruta de dados em informação acionável para gestores e pesquisadores.

As visualizações temporais confirmam e tornam navegáveis os padrões identificados na análise exploratória da primeira etapa. O mapa de calor por hora e dia da semana evidencia a concentração de acidentes em horários de pico e em dias úteis, enquanto os gráficos por ano, por tipo e por bairro permitem comparar regiões e períodos de forma interativa. O mapa georreferenciado, por sua vez, revela a forte concentração espacial na área central da cidade e ao longo dos principais eixos viários.

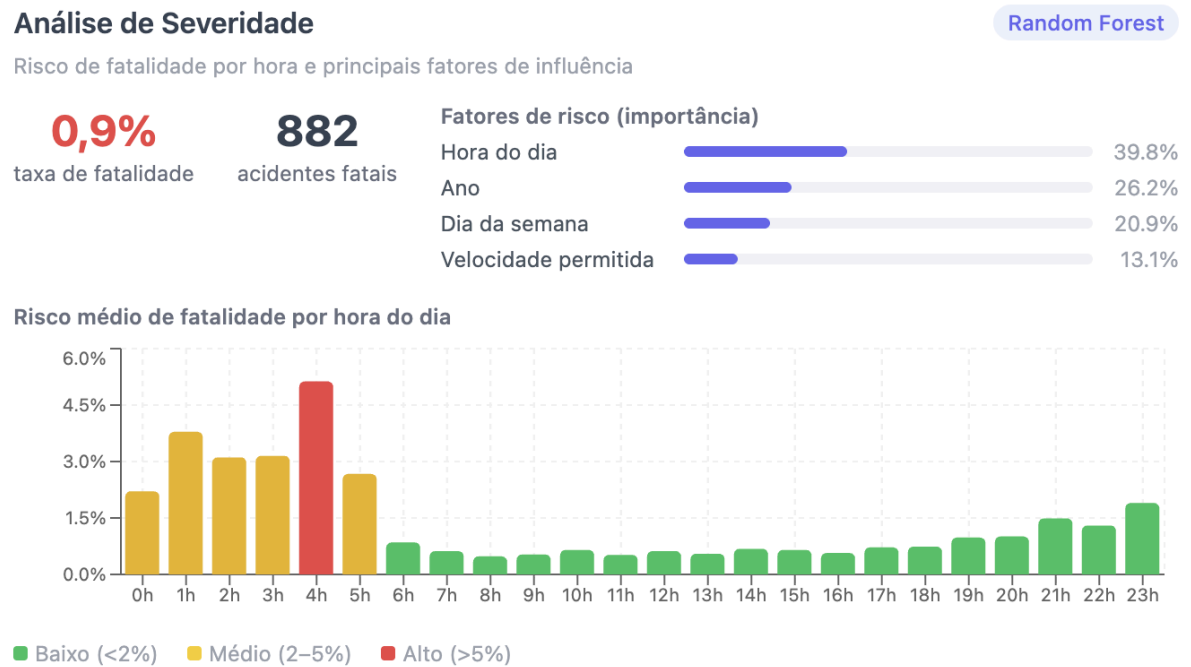
Quanto à dimensão preditiva, o modelo de previsão projeta o número de acidentes para os meses seguintes a partir do comportamento sazonal histórico, acompanhado de uma faixa de incerteza, conforme ilustrado na figura 5. Já a análise de severidade, baseada no modelo Random Forest e ilustrada na figura 6, indica que a hora do dia é o fator mais determinante do risco de fatalidade, respondendo por aproximadamente 39,8% da importância do modelo, seguida pelo ano, com cerca de 26,2%. Esses resultados quantificam, de forma interpretável, a influência de variáveis temporais sobre a gravidade dos sinistros.

Figura 5 - Previsão Naive Sazonal — histórico e projeção dos meses seguintes (± 1 desvio padrão)



Fonte: Elaborado pelo autor (2026).

Figura 6 - Random Forest — importância dos fatores para o risco de fatalidade



Fonte: Elaborado pelo autor (2026).

5. Discussão

Os resultados obtidos demonstram que a passagem da análise exploratória para um sistema web interativo amplia significativamente o alcance dos dados abertos da BHTrans como instrumento de apoio à segurança viária e ao planejamento urbano. Ao reunir, em uma única plataforma, visualizações temporais, espaciais e preditivas com filtros dinâmicos, o sistema permite que diferentes públicos como gestores, pesquisadores, jornalistas e cidadãos explorem padrões reais de quase cem mil acidentes de forma intuitiva, extraindo conhecimento que orienta intervenções direcionadas, como ações de fiscalização, otimização semafórica e campanhas educativas em horários específicos.

A escolha de uma arquitetura desacoplada em três camadas, embora exija um esforço inicial de configuração maior do que soluções monolíticas, mostrou-se adequada aos objetivos do projeto ao favorecer escalabilidade, manutenibilidade, testabilidade e reaproveitamento da API por outros clientes. A integração de competências distintas do curso de Sistemas de Informação: engenharia de software, banco de dados, desenvolvimento web, testes e aprendizado de máquina, conferindo ao trabalho um valor que extrapola a análise de dados, servindo de referência para iniciativas com dados urbanos abertos.

É necessário, contudo, reconhecer limitações. A qualidade das previsões depende da consistência das bases originais e da disponibilidade de variáveis; o modelo preditivo de séries temporais adotado é, deliberadamente, uma linha de base, e modelos mais sofisticados poderiam capturar dinâmicas adicionais. Da mesma forma, a análise de severidade indica associações e importâncias de fatores, mas não estabelece relações causais. Essas limitações apontam caminhos para o refinamento futuro da solução.

6. Conclusão e perspectivas

Esta segunda etapa do trabalho concretizou o objetivo previsto na primeira fase, entregando um sistema web interativo que transforma os dados abertos da BHTrans em uma ferramenta de visualização e análise acessível. Sobre uma arquitetura desacoplada em três camadas, o sistema integra estatísticas agregadas, um mapa georreferenciado e modelos preditivos, consolidando uma visão abrangente dos 97.498 acidentes registrados em Belo Horizonte entre 2016 e 2023 e democratizando o acesso à informação de segurança viária.

Como perspectivas, espera-se o aprimoramento dos modelos preditivos, com a incorporação de técnicas mais avançadas de séries temporais e de classificação; a ampliação das análises georreferenciadas, com a detecção automática de regiões críticas; a publicação

da aplicação para acesso público; e a incorporação de novas fontes e variáveis. Almeja-se que o produto final contribua para uma mobilidade urbana mais eficiente e segura, fortalecendo o uso de dados abertos como instrumento público de transparência e inteligência urbana.

7. Referências

BELO HORIZONTE. Empresa de Transportes e Trânsito de Belo Horizonte – BHTrans. **Dados abertos de acidentes de trânsito em Belo Horizonte**. Disponível em: <https://dados.pbh.gov.br/organization/bhtrans>. Acesso em: 2025.

BANKS, A.; PORCELLO, E. **Learning React**. 2. ed. Sebastopol: O'Reilly Media, 2020.

FIELDING, R. T. **Architectural Styles and the Design of Network-based Software Architectures**. Tese (Doutorado) — University of California, Irvine, 2000.

JEMEROV, D.; ISAKOVA, S. **Kotlin in Action**. Shelter Island: Manning Publications, 2017.

MARTIN, R. C. **Clean Code: A Handbook of Agile Software Craftsmanship**. Upper Saddle River: Prentice Hall, 2008.

WALLS, C. **Spring in Action**. 6. ed. Shelter Island: Manning Publications, 2022.

PROVOST, F.; FAWCETT, T. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. Sebastopol: O'Reilly Media, 2013.