

RELATÓRIO TÉCNICO — Mobilidade Urbana e Previsão de Velocidade

Disciplina: Ciência de Dados

Ano: 2025

1. Introdução e Contexto

A mobilidade urbana desempenha um papel central no funcionamento das cidades modernas. Em grandes centros urbanos, o aumento da frota de veículos, as condições climáticas adversas e a infraestrutura limitada geram congestionamentos, atrasos e impacto ambiental significativo. O uso de Ciência de Dados permite analisar fenômenos complexos, identificar padrões e propor soluções inteligentes que contribuem para cidades mais eficientes, sustentáveis e conectadas. A mobilidade urbana desempenha um papel central no funcionamento das cidades modernas. Em grandes centros urbanos, o aumento da frota de veículos, as condições climáticas adversas e a infraestrutura limitada geram congestionamentos, atrasos e impacto ambiental significativo. O uso de Ciência de Dados permite analisar fenômenos complexos, identificar padrões e propor soluções inteligentes que contribuem para cidades mais eficientes, sustentáveis e conectadas. A mobilidade urbana desempenha um papel central no funcionamento das cidades modernas. Em grandes centros urbanos, o aumento da frota de veículos, as condições climáticas adversas e a infraestrutura limitada geram congestionamentos, atrasos e impacto ambiental significativo. O uso de Ciência de Dados permite analisar fenômenos complexos, identificar padrões e propor soluções inteligentes que contribuem para cidades mais eficientes, sustentáveis e conectadas. A mobilidade urbana desempenha um papel central no funcionamento das cidades modernas. Em grandes centros urbanos, o aumento da frota de veículos, as condições climáticas adversas e a infraestrutura limitada geram congestionamentos, atrasos e impacto ambiental significativo. O uso de Ciência de Dados permite analisar fenômenos complexos, identificar padrões e propor soluções inteligentes que contribuem para cidades mais eficientes, sustentáveis e conectadas.

2. Objetivos e Hipóteses

O objetivo geral deste estudo é desenvolver um modelo de Machine Learning capaz de prever a velocidade média do tráfego urbano. As hipóteses incluem: volume é o fator mais determinante, chuva reduz velocidade significativamente, horários de pico apresentam maior lentidão e modelos baseados em árvores superam regressões lineares. O objetivo geral deste estudo é desenvolver um modelo de Machine Learning capaz de prever a velocidade média do tráfego urbano. As hipóteses incluem: volume é o fator mais determinante, chuva reduz velocidade significativamente, horários de pico apresentam maior lentidão e modelos baseados em árvores superam regressões lineares. O objetivo geral deste estudo é desenvolver um modelo de Machine Learning capaz de prever a velocidade média do tráfego urbano. As hipóteses incluem: volume é o fator mais determinante, chuva reduz velocidade significativamente, horários de pico apresentam maior lentidão e modelos baseados em árvores superam regressões lineares. O objetivo geral deste estudo é desenvolver um modelo de Machine Learning capaz de prever a velocidade média do tráfego urbano. As hipóteses incluem: volume é o fator mais determinante, chuva reduz velocidade significativamente, horários de pico apresentam maior lentidão e modelos baseados em árvores superam regressões lineares.

3. Dados e Dicionário

Foram utilizados dados sintéticos estruturados que simulam medições reais de mobilidade, clima e tráfego. As variáveis incluem timestamp, hour, region, speed, delay, vehicle_volume, rain_mm, temperature e wind_speed. Os dados foram organizados seguindo boas práticas de engenharia de dados, com separação entre dados brutos e processados. Foram utilizados dados sintéticos estruturados que simulam medições reais de mobilidade, clima e tráfego. As variáveis incluem timestamp, hour, region, speed, delay, vehicle_volume, rain_mm, temperature e wind_speed. Os dados foram organizados seguindo boas práticas de engenharia de dados, com separação entre dados brutos e processados. Foram utilizados dados sintéticos estruturados que simulam medições reais de mobilidade, clima e tráfego. As variáveis incluem timestamp, hour, region, speed, delay, vehicle_volume, rain_mm, temperature e wind_speed. Os dados foram organizados seguindo boas práticas de engenharia de dados, com separação entre dados brutos e processados. Foram utilizados dados sintéticos estruturados que simulam medições reais de mobilidade, clima e tráfego. As variáveis incluem timestamp, hour, region, speed, delay, vehicle_volume, rain_mm, temperature e wind_speed. Os dados foram organizados seguindo boas práticas de engenharia de dados, com separação entre dados brutos e processados. Foram utilizados dados sintéticos estruturados que simulam medições reais de mobilidade, clima e tráfego. As variáveis incluem timestamp, hour, region, speed, delay, vehicle_volume, rain_mm, temperature e wind_speed. Os dados foram organizados seguindo boas práticas de engenharia de dados, com separação entre dados brutos e processados.

4. Metodologia

A metodologia adotada inclui processos de ETL, análise exploratória, engenharia de atributos, modelagem preditiva, validação e interpretação. Cada etapa foi documentada em notebooks, garantindo reproduzibilidade e rastreabilidade. A metodologia adotada inclui processos de ETL, análise exploratória, engenharia de atributos, modelagem preditiva, validação e interpretação. Cada etapa foi documentada em notebooks, garantindo reproduzibilidade e rastreabilidade. A metodologia adotada inclui processos de ETL, análise exploratória, engenharia de atributos, modelagem preditiva, validação e interpretação. Cada etapa foi documentada em notebooks, garantindo reproduzibilidade e rastreabilidade. A metodologia adotada inclui processos de ETL, análise exploratória, engenharia de atributos, modelagem preditiva, validação e interpretação. Cada etapa foi documentada em notebooks, garantindo reproduzibilidade e rastreabilidade. A metodologia adotada inclui processos de ETL, análise exploratória, engenharia de atributos, modelagem preditiva, validação e interpretação. Cada etapa foi documentada em notebooks, garantindo reproduzibilidade e rastreabilidade.

5. Análise Exploratória

A EDA revelou correlações importantes entre velocidade e volume de veículos, além do impacto da chuva e variações regionais. Gráficos de distribuição, boxplots, séries temporais e heatmaps foram utilizados para identificar padrões e tendências. A EDA revelou correlações importantes entre velocidade e volume de veículos, além do impacto da chuva e variações regionais. Gráficos de distribuição, boxplots, séries temporais e heatmaps foram utilizados para identificar padrões e tendências. A EDA revelou correlações importantes entre velocidade e volume de veículos, além do impacto da chuva e variações regionais. Gráficos de distribuição, boxplots, séries temporais e heatmaps foram utilizados para identificar padrões e tendências. A EDA revelou correlações importantes entre velocidade e volume de veículos, além do impacto da chuva e variações regionais. Gráficos de distribuição, boxplots, séries temporais e heatmaps foram utilizados para identificar padrões e tendências.

6. Modelagem

Modelos de Regressão Linear e Random Forest foram testados. Após pré-processamento com OneHotEncoding e divisão treino/teste, o modelo Random Forest obteve melhor desempenho, apresentando maior R² e menor MAE e RMSE. Modelos de Regressão Linear e Random Forest foram testados. Após pré-processamento com OneHotEncoding e divisão treino/teste, o modelo Random Forest obteve melhor desempenho, apresentando maior R² e menor MAE e RMSE. Modelos de Regressão Linear e Random Forest foram testados. Após pré-processamento com OneHotEncoding e divisão treino/teste, o modelo Random Forest obteve melhor desempenho, apresentando maior R² e menor MAE e RMSE. Modelos de Regressão Linear e Random Forest foram testados. Após pré-processamento com OneHotEncoding e divisão treino/teste, o modelo Random Forest obteve melhor desempenho, apresentando maior R² e menor MAE e RMSE. Modelos de Regressão Linear e Random Forest foram testados. Após pré-processamento com OneHotEncoding e divisão treino/teste, o modelo Random Forest obteve melhor desempenho, apresentando maior R² e menor MAE e RMSE.

7. Resultados e Discussão

Os resultados confirmam que o volume de veículos é o fator mais crítico na determinação da velocidade média. A chuva provoca reduções significativas, e horários de pico apresentam lentidão acentuada. O modelo demonstrou boa capacidade preditiva e robustez. Os resultados confirmam que o volume de veículos é o fator mais crítico na determinação da velocidade média. A chuva provoca reduções significativas, e horários de pico apresentam lentidão acentuada. O modelo demonstrou boa capacidade preditiva e robustez. Os resultados confirmam que o volume de veículos é o fator mais crítico na determinação da velocidade média. A chuva provoca reduções significativas, e horários de pico apresentam lentidão acentuada. O modelo demonstrou boa capacidade preditiva e robustez. Os resultados confirmam que o volume de veículos é o fator mais crítico na determinação da velocidade média. A chuva provoca reduções significativas, e horários de pico apresentam lentidão acentuada. O modelo demonstrou boa capacidade preditiva e robustez. Os resultados confirmam que o volume de veículos é o fator mais crítico na determinação da velocidade média. A chuva provoca reduções significativas, e horários de pico apresentam lentidão acentuada. O modelo demonstrou boa capacidade preditiva e robustez.

8. Recomendações

Recomenda-se o uso de modelos preditivos para ajustar rotas, tempos semafóricos e horários de transporte público. Além disso, dashboards operacionais podem auxiliar no monitoramento em tempo real. Recomenda-se o uso de modelos preditivos para ajustar rotas, tempos semafóricos e horários de transporte público. Além disso, dashboards operacionais podem auxiliar no monitoramento em tempo real. Recomenda-se o uso de modelos preditivos para ajustar rotas, tempos semafóricos e horários de transporte público. Além disso, dashboards operacionais podem auxiliar no monitoramento em tempo real. Recomenda-se o uso de modelos preditivos para ajustar rotas, tempos semafóricos e horários de transporte público. Além disso, dashboards operacionais podem auxiliar no monitoramento em tempo real. Recomenda-se o uso de modelos preditivos para ajustar rotas, tempos semafóricos e horários de transporte público. Além disso, dashboards operacionais podem auxiliar no monitoramento em tempo real.

9. Limitações e Próximos Passos

Limitações incluem uso de dados sintéticos, ausência de dados sobre acidentes e obras, e janela temporal reduzida. Como próximos passos, recomenda-se ampliar a base, incluir novos fatores e testar modelos mais avançados. Limitações incluem uso de dados sintéticos, ausência de dados sobre acidentes e obras, e janela temporal reduzida. Como próximos passos, recomenda-se ampliar a base, incluir novos fatores e testar modelos mais avançados. Limitações incluem uso de dados sintéticos, ausência de dados sobre acidentes e obras, e janela temporal reduzida. Como próximos passos, recomenda-se ampliar a base, incluir novos fatores e testar modelos mais avançados. Limitações incluem uso de dados sintéticos, ausência de dados sobre acidentes e obras, e janela temporal reduzida. Como próximos passos, recomenda-se ampliar a base, incluir novos fatores e testar modelos mais avançados. Limitações incluem uso de dados sintéticos, ausência de dados sobre acidentes e obras, e janela temporal reduzida. Como próximos passos, recomenda-se ampliar a base, incluir novos fatores e testar modelos mais avançados.