

Previsão do Uso de Praças de Pedágio: Uma Análise Baseada em Machine Learning

Anderson Mendes¹, Elias Neto¹, Vitor Rafael¹

¹Faculdade Nova Roma - Av. Adjar da Silva Casé, 800 - Indianópolis, Caruaru - PE, 55024-740

Resumo. *O aumento das frotas de veículos nos centros urbanos brasileiros demanda a modernização dos sistemas de infraestrutura rodoviária. Este estudo explora modelos preditivos de inteligência artificial para estimar o volume de tráfego em praças de pedágio, utilizando dados da ANTT e a metodologia CRISP-DM. Foram abordados modelos ARMA, ARIMA e SARIMA, avaliados por métricas como AIC, BIC, MAE e MSE. Após otimização de hiperparâmetros, os modelos de ML proporcionaram previsões viáveis para o fluxo de veículos. Conclui-se que o ML pode ser uma ferramenta valiosa para a gestão eficiente do tráfego rodoviário.*

Abstract. *The increase in vehicle fleets in Brazilian urban centers demands the modernization of road infrastructure systems. This study explores artificial intelligence predictive models to estimate the volume of traffic at toll plazas, using data from ANTT and the CRISP-DM methodology. ARMA, ARIMA and SARIMA models were covered, evaluated by metrics such as AIC, BIC, MAE and MSE. After hyperparameter optimization, the ML models provided viable predictions for vehicle flow. It is concluded that ML can be a valuable tool for the efficient management of road traffic.*

1. Introdução

É notório que com o passar dos anos as frotas de veículos vem cada vez mais aumentando nas cidades e em grandes centro urbanos brasileiros em virtude de fatores sociais e econômicos como destacado por [Ramis and Santos 2012]

“Aliada a uma postura cultural histórica das classes médias que coloca a posse do automóvel como símbolo de status, de absoluta mobilidade e “liberdade”, certamente aumentam suas vendas e, conseqüentemente, sua presença nas ruas.”

Isso por sua vez faz com que seja crescente a necessidade de que os sistemas que fornecem infraestrutura e suporte para essa frota sejam cada vez mais otimizados e modernizados, para que possam suportar todas esta demanda.

Para aprimorar a eficiência na gestão do tráfego em rodovias e praças de pedágio, é essencial considerar a previsão precisa do volume de veículos em diferentes momentos e locais, levando em conta fatores como condições climáticas, eventos especiais e feriados [Yang et al. 2012]. A implementação de tecnologias como a cobrança em rodovias de pedágios abertos (RPA) tem demonstrado melhorias nas operações de tráfego em praças de pedágio, reduzindo a exposição dos motoristas a acidentes e beneficiando a segurança [Gordin et al. 2011].

A previsão precisa do fluxo de tráfego tem implicações profundas[Neri and Silva Júnior 2023]. Ela pode melhorar a segurança nas estradas, otimizar a gestão do tráfego, reduzir o tempo de viagem e até mesmo contribuir para a redução das emissões de carbono[Barth and Boriboonsomsin 2008]. As autoridades e planejadores podem utilizar estas previsões para tomar decisões informadas e implementar estratégias eficazes, resultando em benefícios abrangentes para a sociedade e o meio ambiente.

Aqui é onde a aprendizagem de máquina (Machine Learning - ML) pode desempenhar um papel crucial. Com a sua capacidade de identificar padrões complexos e tendências nos dados históricos, o ML pode ser usado para desenvolver modelos preditivos para estimar o volume de tráfego em uma determinada praça de pedágio. Esses modelos podem levar em consideração uma variedade de fatores, incluindo mas não se limitando a condições climáticas, eventos especiais e feriados.

Ao superar esse desafio, podemos fornecer às autoridades de trânsito e operadores de rodovias uma ferramenta valiosa para o planejamento e gestão eficazes do tráfego. Isso, por sua vez, pode levar a uma melhoria significativa na segurança e eficiência do transporte rodoviário.

Diante do exposto acima, o objetivo deste estudo é explorar a aplicação de modelos preditivos baseados em inteligência artificial para estimar o volume de tráfego em praças de pedágio, de modo a contribuir para o aprimoramento da gestão do tráfego em rodovias, fornecendo às autoridades responsáveis informações que possam subsidiar a tomada de decisões e otimizar a operação das praças de pedágio. Para isso o nosso estudo será paltado na base de dados fornecida pela ANTT a partir da qual implementamos a metodologia do CRISP DM para a condução de nossa pesquisa.

Vale ressaltar que não temos a inteção de implementar dashboards automáticos para apresentação de métricas,desejamos proporcionar insights para uma melhor gestão do tráfego rodoviário.

2. Fundamentação teórica

2.1. Análise do Tráfego

De acordo com o *Plano Nacional de Logística 2035*[da Infraestrutura 2021], até o ano de 2017, o transporte interurbano de pessoas possuía, em seus três primeiros lugares, o transporte rodoviário privado com 57,66%, o transporte aeroviário com 24,75% e, em seguida, o transporte rodoviário por ônibus com 16,83% dos transportes. Logo, as rodovias são usadas em 74,49% dos deslocamentos interurbanos de pessoas. Ainda de acordo com o *Plano Nacional de Logística 2035*, temos que, no ano de 2017, o transporte rodoviário foi responsável pela movimentação de 68,54% das cargas em nosso país. Diante desses números, fica clara a importância que as rodovias exercem em nosso país e, portanto, é imprescindível que sejam feitos investimentos que busquem otimizar as condições para que o tráfego possa fluir da melhor forma possível.

2.2. Aprendizado de Máquina (Machine Learning - ML)

De acordo com *Monard e Baranauskas*[Monard and Baranauskas 2003], "Aprendizado de Máquina é uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado, bem como a construção de sistemas capazes de adquirir

conhecimento de forma automática". Dessa forma, em nosso estudo, usamos três técnicas de *Machine Learning* para que pudéssemos verificar com qual delas poderíamos chegar a um melhor grau de confiabilidade em nossas previsões. A saber Arma, Arima e Sarima sobre as quais discorreremos melhor nas seções seguintes.

De acordo com Gujarati(2000) apud Silva [SILVA] "Os modelos Auto Regressivos (AR) e Média Móvel (MA) são baseados na suposição de que a série temporal seja gerada através de um sistema linear, e que possuem um termo de erro aleatório não correlacionado, com média zero e variância constante". Assim o modelo de aprendizado de máquina **ARMA** visa fazer a previsão de séries temporais estacionárias.

Já o modelo de aprendizado de máquina **ARIMA** de acordo com Santos[SANTOS JÚNIOR et al. 2022] é "uma extensão do ARMA e possui a capacidade de realizar diferenças sucessivas nos dados, removendo a tendência e tornando uma série não-estacionária em estacionária."

E por sua vez o modelo **SARIMA** é uma extensão do ARIMA que leva em conta a sazonalidade. Ele adiciona termos sazonais AR, MA e de diferenciação ao modelo ARIMA. Isso permite que o modelo capture padrões que se repetem em intervalos regulares.

Para avaliar a performance desses modelos, utilizamos diversas métricas estatísticas, entre elas AIC, BIC, MAE e MSE. Abaixo, explicamos cada uma dessas métricas:

- **AIC (Akaike Information Criterion):** AIC é uma medida utilizada para comparar a qualidade de diferentes modelos estatísticos para um conjunto de dados. A fórmula para o AIC é: $AIC = 2k - 2\ln(L)$, onde k é o número de parâmetros do modelo e L é a função de verossimilhança máxima do modelo. Modelos com menor valor de AIC são preferíveis, pois indicam um melhor ajuste ao conjunto de dados com menor complexidade.
- **BIC (Bayesian Information Criterion):** Semelhante ao AIC, o BIC também é usado para comparação de modelos, mas penaliza mais fortemente a complexidade do modelo. A fórmula do BIC é: $BIC = k\ln(n) - 2\ln(L)$, onde n é o número de observações. Assim como no AIC, modelos com menor BIC são preferíveis.
- **MAE (Mean Absolute Error):** MAE é a média das diferenças absolutas entre os valores previstos e os valores observados. A fórmula é: $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$, onde y_i são os valores observados e \hat{y}_i são os valores previstos. Um MAE menor indica previsões mais precisas.
- **MSE (Mean Squared Error):** MSE é a média das diferenças quadradas entre os valores previstos e os valores observados. A fórmula é: $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$. MSE é mais sensível a grandes erros devido ao termo quadrático, fazendo com que um MSE menor indique previsões mais precisas.

Essas métricas nos permitem avaliar e comparar a precisão e a eficiência dos modelos de aprendizado de máquina aplicados às nossas séries temporais, auxiliando na escolha do modelo mais adequado para as previsões.

2.3. Trabalhos Relacionados

RADAR[Mendonça 2022] é um mecanismo de previsão de fluxo de tráfego de veículos em vias urbanas, baseado em Aprendizado Federado. O autor destaca a dificuldade

de gerar um modelo abrangente que possa prever resultados em múltiplas vias, devido às particularidades associadas a cada uma delas. Para contornar esse obstáculo, é proposto o uso do modelo de aprendizado de máquina Federated Learning (FL), permitindo a integração de múltiplos pontos de entrada para a construção de um único modelo. Na conclusão do trabalho, é apresentado que o modelo proposto, chamado de RADAR, foi competitivo com os modelos centralizados, com uma leve melhoria nas métricas de erro, especialmente em conjuntos de dados maiores.

Os autores de "Deep Temporal Convolutional Networks for Short-Term Traffic Flow Forecasting"[Zhao et al. 2019] começam destacando os benefícios e melhorias que um sistema inteligente de controle de tráfego pode trazer para a construção de uma cidade inteligente. O escopo do projeto é definido, focando na previsão de curtos espaços de tempo para que congestionamentos possam ser previstos e medidas possam ser tomadas para minimizá-los ou evitá-los. As principais técnicas de aprendizado de máquina utilizadas incluem Convolutional Neural Networks (CNN) e Long short-term memory (LSTM). Os resultados indicam que o modelo tem uma taxa de previsão com até 95% de precisão, mostrando-se uma ferramenta importante para previsão de tráfego de curto prazo nas cidades.

No estudo sobre o aperfeiçoamento do modelo de previsão de conflitos em rodovias através de Deep Learning [Caleffi and Cybis], são elencados diversos fatores relevantes para o acontecimento de eventos que geram conflitos, sendo citada a relevância do fluxo de veículos. O autor emprega técnicas de Deep Learning (redes neurais), alcançando uma precisão que varia de 76,28% a 85,92% após a otimização dos hiperparâmetros.

3. Metodologia

Para a condução de nossa pesquisa fizemos o uso da metodologia **CRISP-DM** que consiste em uma metodologia dividida em seis etapas, sendo elas:

- Compreensão do negócio.
- Entendimento dos dados.
- Preparação dos dados.
- Modelagem.
- Avaliação.
- Implementação.

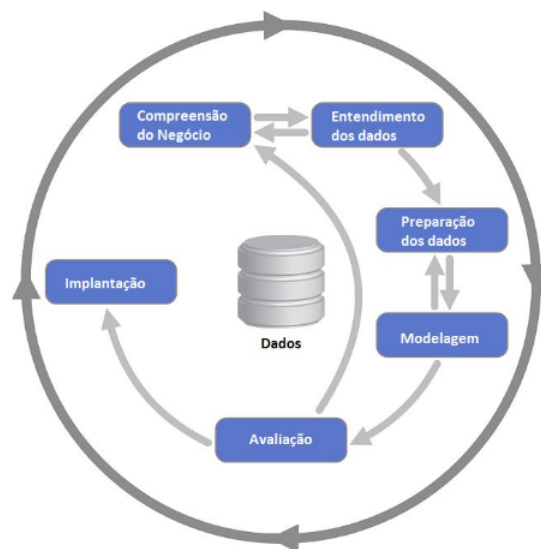


Figura 1. Fases do CRISP-DM. Fonte: [Shearer, 2000]

Como podemos observar na imagem a sequência das fases não é rígida e pode ser intercambiável permitindo assim que o processo ocorra de forma simples, e cada fase busca avançar na correta apropriação dos dados. Na primeira fase o foco principal está em entender quais os objetivos do projeto a partir da perspectiva do negócio. Em seguida na fase de **entendimento dos dados** a atenção passa agora a ser voltada para que seja possível um olhar mais atento aos dados disponíveis e seu valor para que seja alcançado o objetivo traçado na etapa anterior. Na **preparação dos dados** temos uma das fases que mais demanda tempo, pois é nela que devemos implementar todo o tipo de tratamento requerido pela base de dados removendo vícios, valores incoerentes, nulos e realizando os ajustes necessários para que a modelagem possa acontecer. É na **modelagem** em que os algoritmos de ML serão executados e serão feitas diversas iterações visando obter os melhores hiper-parâmetros em cada uma das técnicas utilizadas bem como serão testadas quais técnicas possuem um melhor desempenho na solução do problema proposto. Na fase de **Avaliação** devemos analisar se com as técnicas utilizadas até então os objetivos que foram traçados estão sendo atingidos de forma satisfatória. E por fim na fase de **Implantação** os novos conhecimentos descobertos são usados para proporcionar melhorias a situação abordada.

Desta feita passaremos a partir de agora a implementar essa metodologia para a problemática por nós abordada. Como elencado em nossa introdução o nosso objetivo é construir um modelo que permita a previsão do fluxo de veículos em determinada via a partir dos dados obtidos em períodos anteriores.

Uma vez em posse dos dados divulgados pela ANTT, iniciaremos o pré-processamento dos mesmos, fase em que estamos interessados em melhorar a qualidade das informações que dispomos, removendo inconsistências, dados irrelevantes para nosso estudo e valores desconhecidos ou incoerentes. Estas ações têm como objetivo preparar os dados para a fase seguinte. Assim estaremos realizando tanto o entendimento quanto a preparação dos dados.

Na primeira carga dos dados, que consiste em dados de janeiro de 2011 a abril de

2024, tivemos um total de 1.434.895 linhas e 7 colunas, descritas no dicionário de dados (tabela 1), possuem as seguintes descrições:

Tabela 1. Dicionário de dados

| Campo | Descrição |
|-----------------|---|
| Concessionaria | Nome da concessionária que prestou as informações |
| mes_ano | Data de referência com dados da praça de pedágio. Ex.: jan-17 |
| sentido | Orientação de direção geral. Ex.: norte, sul, leste, oeste, sentido 1, sentido 2 |
| praca | Nome dado para a praça de pedágio. Ex.: P1, Praça 02, BR-040/RJ Km 45,50 |
| categoria | Representação veicular na tabela de tarifas. Ex.: Categoria 1; Veículo de passeio 2 eixos; Veículo comercial 3 eixos. |
| tipo_de_veiculo | Classificação veicular. Ex.: comercial, moto, passeio |
| volume_total | Volume total de veículos que passaram na praça informada. Ex.: 14860 |

Além das colunas descritas no dicionário de dados, existe ainda a coluna 'tipo_cobrança', que aparentemente informa se a cobrança foi feita de forma automática ou não.

Sendo assim, as medidas adotadas em nosso pré-processamento foram:

1. Primeiramente aplicamos a remoção de linhas duplicadas, gerando a remoção de 99.286 restando portanto agora 1.335.609 linhas.
2. Logo em seguida foram removidas de nossos dados as colunas que julgamos ser desnecessárias ao nosso caso de estudo. São elas:
 - **Concessionaria** - A informação trazida por esta coluna nos diz apenas qual a empresa responsável pela administração de determinada praça e pedágio o que não é relevante para nossa análise.
 - **Tipo_cobrança** - A informação trazida por esta coluna é apenas sobre qual a taxa de pagamento para o veículo em questão, que não é relevante para a nossa análise.
 - **categoria** - A informação trazida por esta coluna está fortemente ligada a coluna anterior, fazendo a classificação de um veículo em categorias de acordo com o seu tipo, e portanto não é relevante para a nossa análise.
3. A seguir foi realizada a uniformização dos dados constantes na coluna de datas, visto que haviam múltiplos formatos gerando assim erros em nosso processamento. Apesar disso não houve modificação na quantidade de linhas.
4. Depois foram realizados ajustes na coluna **volume_total** para que todos os valores estivessem em formato inteiro permitindo assim sua manipulação numérica.
5. Posteriormente, foi necessário também que fosse ajustada a capitalização das fontes das colunas **sentido**, **praca** e **tipo_de_veiculo**, que eram disformes, trazendo ora caracteres maiúsculos e ora minúsculos, inviabilizando assim que houvesse uma filtragem de forma efetiva.
6. Logo após, foi realizado o agrupamento das informações de veículos que, apesar de estarem classificados como categorias de cobrança diferentes, são do mesmo tipo de veículo, aqui classificados como **Passeio**, **Moto** ou **Comercial**. Além

disso, foi ajustada também a granularidade da base de dados, pois a partir do ano de 2024 a granularidade é trazida em dias enquanto, nos demais anos, é apenas de meses, sendo necessário assim que os dados diários sejam agrupados nos respectivos meses. Após este agrupamento, ficamos com um total de 98.019 linhas em nosso dataframe.

Dessa forma concluímos as etapas necessárias para que tivéssemos uma base de dados na qual fosse possível empregar técnicas para o processamento das informações. E para fins de visualização temos a tabela 2 com a atual forma de nossos dados até o momento.

Tabela 2. Exemplos de Dados

| | mes_ano | sentido | praça | tipo_de_veiculo | volume_total |
|-------|------------|------------|--|-----------------|--------------|
| 18082 | 2015-03-01 | crescente | praça 02 br-324/ba km 551,094 sentido 1: leste | moto | 8264 |
| 51961 | 2023-02-01 | decrecente | p7 - jataí br 364/go km 156,550 | passeio | 12163 |
| 83768 | 2012-04-01 | decrecente | praça 06 br-116/pr km 57,10 | passeio | 106364 |
| 49815 | 2021-11-01 | decrecente | p4 | moto | 7670 |
| 41148 | 2021-06-01 | crescente | praça 07 br-381/mg km 596,935 | passeio | 104017 |

Isso nos permitiu chegar a um total de 344 praças de pedágio, passaremos a partir deste ponto a buscar qual será a melhor praça de pedágio para que possamos treinar e testar nossos modelos. Com esta finalidade criamos a tabela 3 na qual é exibida de forma decrescente as praças de pedágio que possuíam a maior quantidade de entradas, para que fosse selecionada aquela onde os modelos obtivessem a maior quantidade de dados possível para realizar o seu treinamento.

Tabela 3. Frequência de Praças de Pedágio

| | Praça | Count |
|---|------------------------------|-------|
| 0 | praça 03 br-040/mg km 816,70 | 960 |
| 1 | praça 01 br-040/rj km 102,00 | 960 |
| 2 | praça 02 br-040/rj km 45,50 | 960 |
| 3 | praça 05 br-116/sp km 542,90 | 936 |
| 4 | praça 03 br-101/sc km 79,30 | 936 |

Observa-se portanto que as três primeiras praças de nossa base de dados possuem todas 960 ocorrências, para fins de nosso estudo iremos utilizar **praça 03 br-040/mg km 816,70**. É importante também que sejam especificados o tipo de veículo e o sentido para gerar uma maior assertividade em nosso modelo. Especificaremos portanto o tipo de veiculo como sendo *Comercial* e o sentido da via com *Crescente*.

Ao aplicar esses filtros, chegamos a um total de 148 linhas de dados, abrangendo o período de janeiro de 2011 a abril de 2024. Além disso, obtivemos informações relevantes, como pode ser observado na tabela 4.

Tabela 4. Dados do volume de tráfego

| | |
|-------------------------------|----------|
| Fluxo Médio | 41566.56 |
| Menor Fluxo Registrado | 26527 |
| Maior Fluxo Registrado | 56152 |

A partir de então fizemos a separação de nossa base de dados em duas partes que serão usadas para treino e teste dos modelos. O ponto usado para fazer essa separação foi o mês de junho de 2023.

Após as etapas elencadas até aqui foi possível que gerássemos os gráficos das figuras 2, 3 nos quais podemos observar a distribuição do fluxo de veículos comerciais na praça de pedágio 03 br-040/mg km 816,70 no sentido crescente.

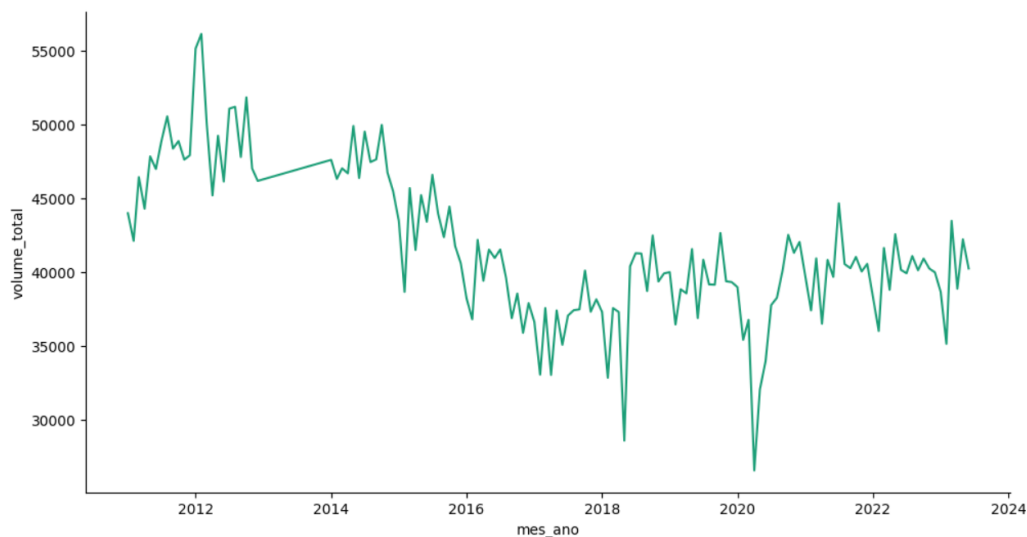


Figura 2. Fluxo da praça 03 br-040/mg km 816,70(contínuo). Fonte: próprio autor

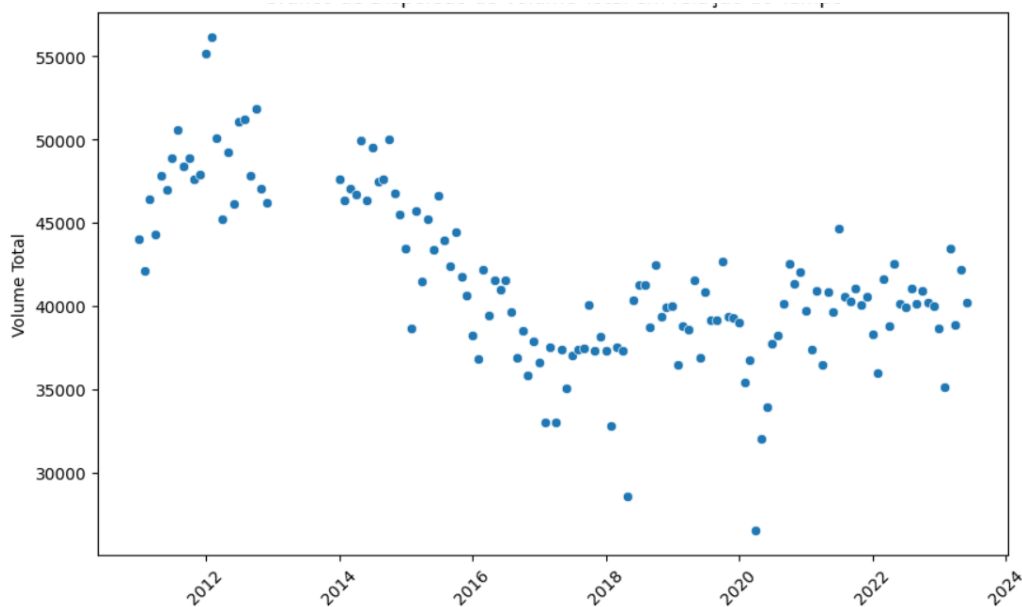


Figura 3. Fluxo da praça 03 br-040/mg km 816,70 (discreto). Fonte: próprio autor

Logo em seguida, identificamos e removemos os outliers presentes em nosso conjunto de informações, visto que, apesar de serem informações reais, sua presença no con-

junto de dados durante a execução do modelo poderia gerar vieses. Chegamos assim ao gráfico da Figura 4, onde os outliers foram identificados antes de sua remoção.

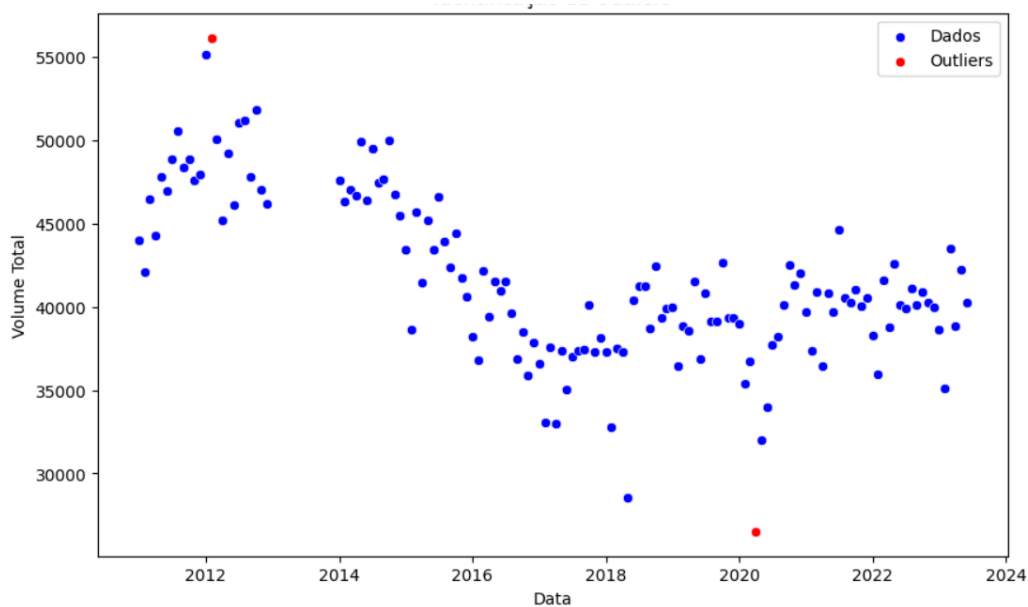


Figura 4. Outliers. Fonte: próprio autor

Com isso finalizamos todas as etapas necessárias para que os dados estivessem prontos para o emprego dos modelos de Machine Learning.

4. Resultados

O primeiro modelo executado para nosso estudo foi o ARMA com os parâmetros padrão onde p , d , e q foram respectivamente 2, 0 e 2, gerando o modelo observado na figura 5

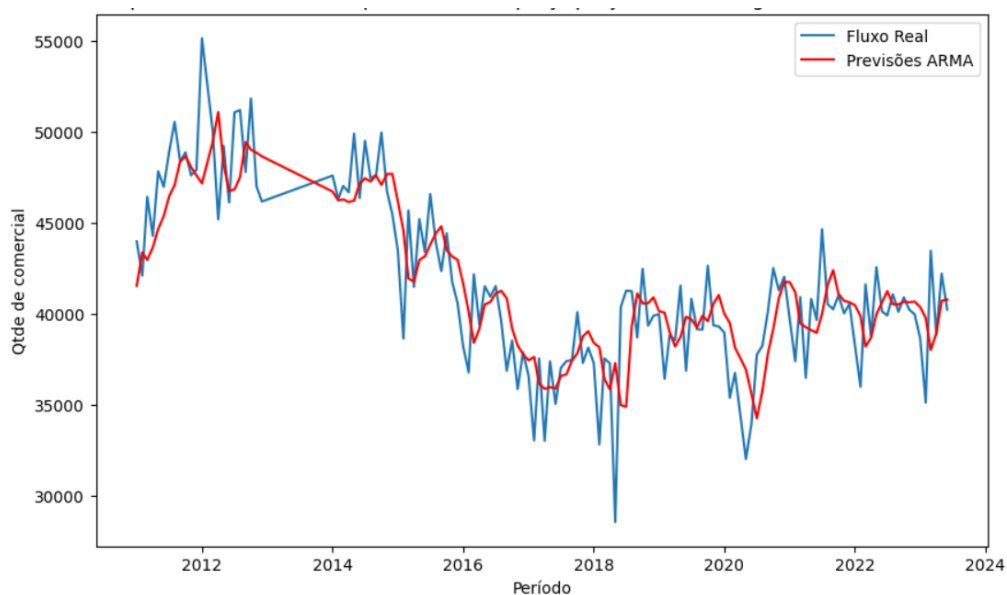


Figura 5. MODELO ARMA (padrão). Fonte: próprio autor

Logo em seguida fizemos o processo de otimização de hiper parâmetros Random Search no qual definimos os seguintes ranges de busca para os parâmetros $p(0,8)$, d fixo em 0 e $q(0, 8)$. Após este processo de otimização chegamos aos parâmetros de $p=7$, $d=0$ e $q=4$. gerando o modelo observado na figura 6

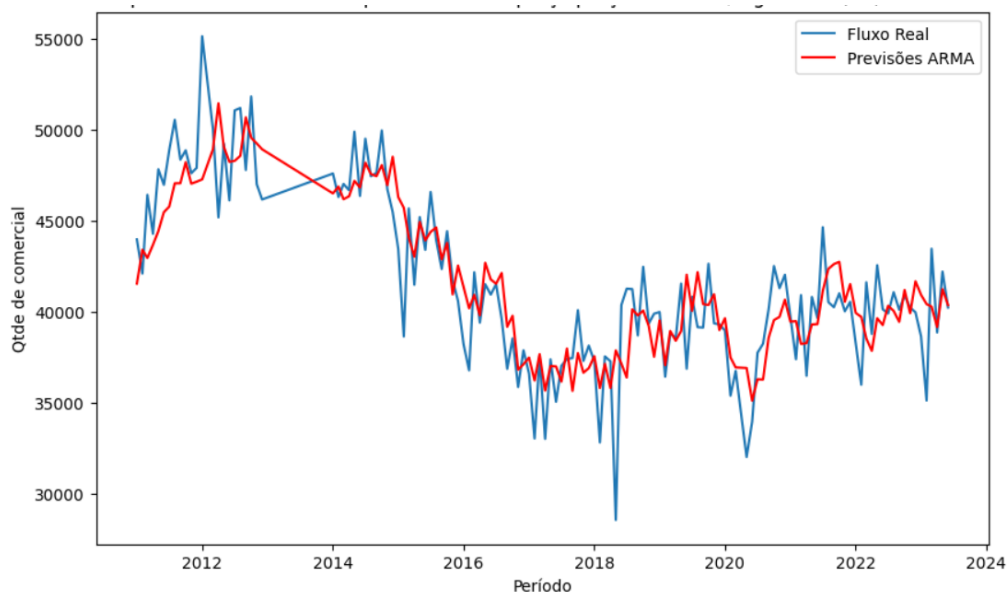


Figura 6. MODELO ARMA (otimizado). Fonte: próprio autor

Foi executado também o modelo ARIMA com os parâmetros padrão onde $p=2$, $d=4$ e $q=2$, gerando o modelo observado na figura 7

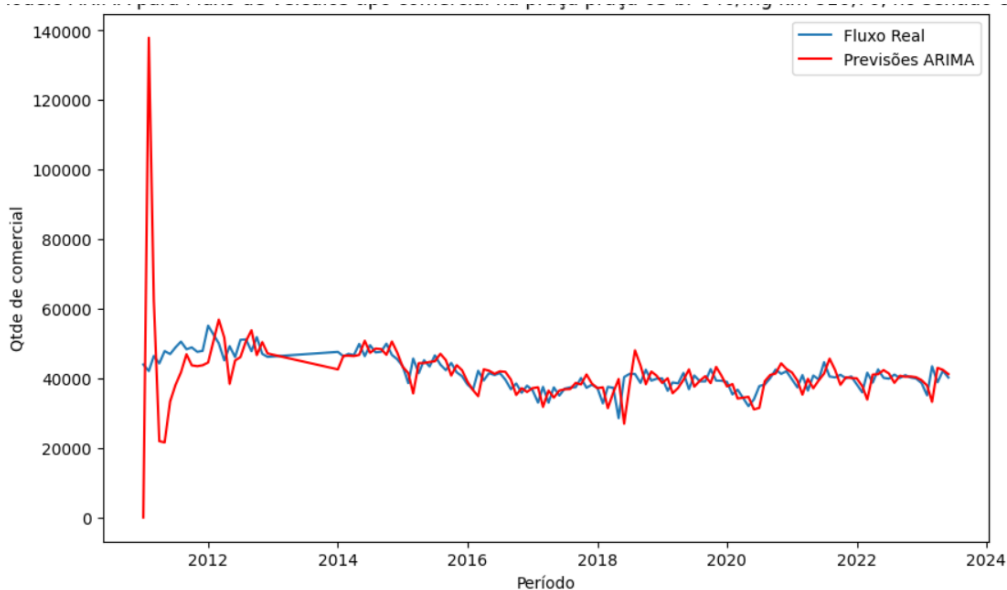


Figura 7. MODELO ARIMA (padrão). Fonte: próprio autor

Após a otimização dos parâmetros do ARIMA com o Random Search obtivemos

$p=9$, $d=0$ e $q=2$, foram usados os intervalos de variação de 0 a 10 para todos os parâmetros. Dessa forma obtivemos o modelo observado na figura 8

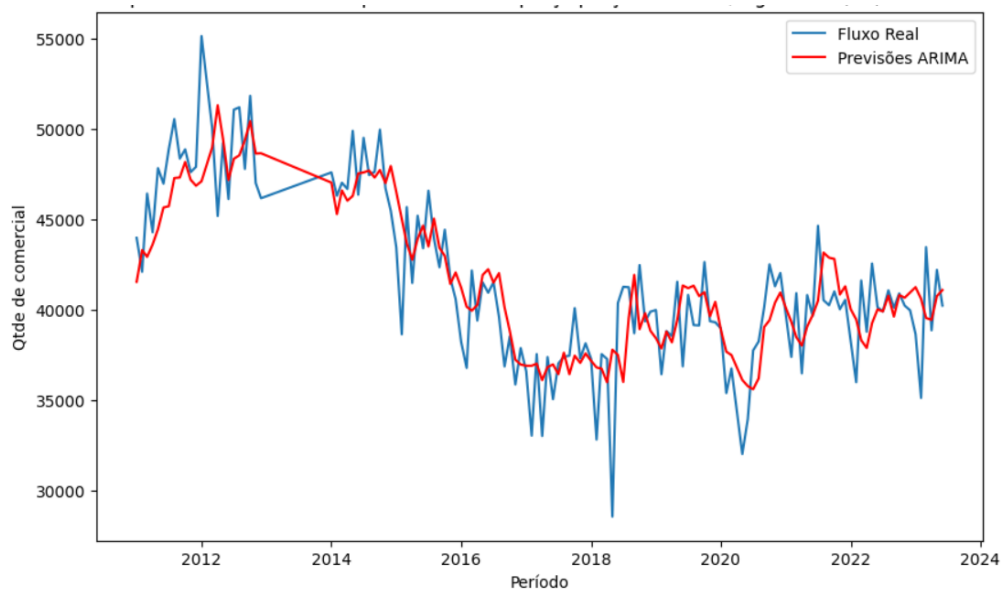


Figura 8. MODELO ARIMA (otimizado). Fonte: próprio autor

Executamos ainda o modelo SARIMA com os parâmetros padrão, a saber $p=1$, $d=1$, $q=1$, $P=1$, $D=1$, $Q=1$ e $S=10$, gerando o modelo observado na figura 9

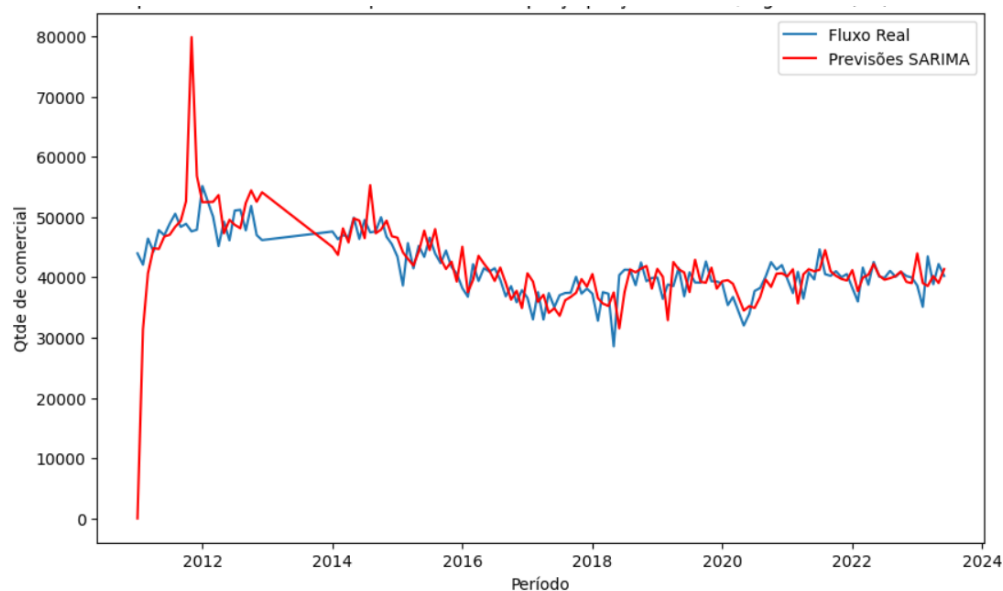


Figura 9. MODELO SARIMA (padrão). Fonte: próprio autor

Geramos também uma otimização de algoritmos por meio de Random Search, na qual os parâmetros p , q , P e Q poderiam variar entre 0-3, os parâmetros d e D poderiam variar entre 0-2 e o parâmetro S variava entre 10-12. Dessa forma, o modelo que obteve

melhor desempenho usou os valores $p=1$, $d=0$, $q=1$, $P=1$, $D=1$, $Q=2$ e $S=12$, gerando o modelo observado na figura 10

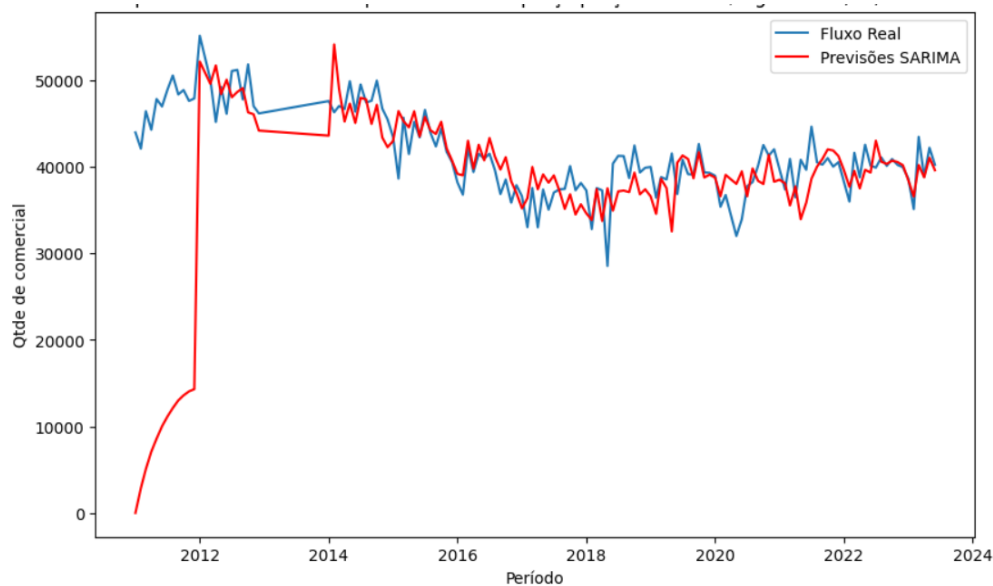


Figura 10. MODELO SARIMA (otimizado). Fonte: próprio autor

Para verificarmos se houve melhoria entre os modelos antes e depois de suas respectivas otimizações, foram utilizadas as métricas AIC e BIC que são observadas nas tabelas 5 e 6.

| Tabela 5. Modelos Padrão | | |
|--------------------------|-----------|-----------|
| Modelo | AIC | BIC |
| ARMA | 2555.2266 | 2575.6152 |
| ARIMA | 2678.9696 | 2701.9099 |
| SARIMA | 2381.4160 | 2401.2700 |

| Tabela 6. Modelos Otimizados | | |
|------------------------------|-----------|-----------|
| Modelo | AIC | BIC |
| ARMA | 2524.6586 | 2562.5232 |
| ARIMA | 2529.5706 | 2567.4351 |
| SARIMA | 2342.0337 | 2358.9554 |

Podemos observar que em todos os modelos usados após a otimização feita usando Random Search houve uma melhora dos indicadores. Utilizamos em seguida os modelos otimizados para que fossem gerados os valores correspondentes as datas que foram separadas para teste para uma segunda comparação utilizando agora as métricas de Erro médio absoluto(MAE) e o erro quadrático médio (MSE), obtendo os valores que constam na tabela 7 e que podem ser observados no gráfico da figura 11.

Tabela 7. Desempenho dos Modelos

| Modelo | MAE | MSE |
|--------|-----------|--------------|
| ARMA | 2718.3295 | 9599486.9629 |
| ARIMA | 2684.8655 | 8739785.7833 |
| SARIMA | 2138.9664 | 5958307.6461 |

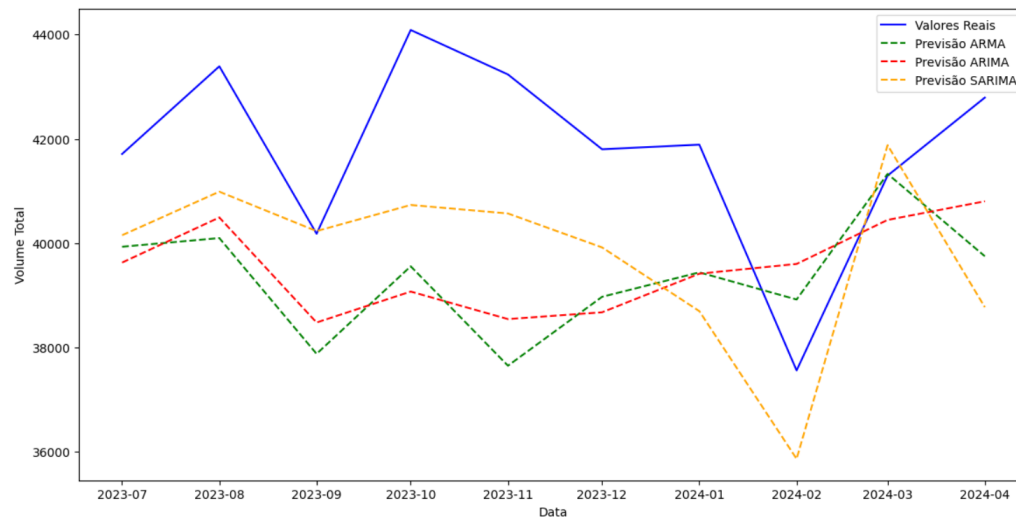


Figura 11. Comparação entre modelos. Fonte: próprio autor

Observamos que em todas as métricas utilizadas o modelo **SARIMA** obteve o melhor desempenho e portanto decidimos utilizá-lo para que fossem feitas as previsões de fluxo de veículos para o ano de 2024. Para isso, fixamos os valores encontrados pela otimização feita e realizamos um novo treinamento com o modelo utilizando agora toda a base de dados disponível visando obter o melhor modelo possível. As previsões são apresentadas no gráfico da figura 12 bem como sua margem de erro.

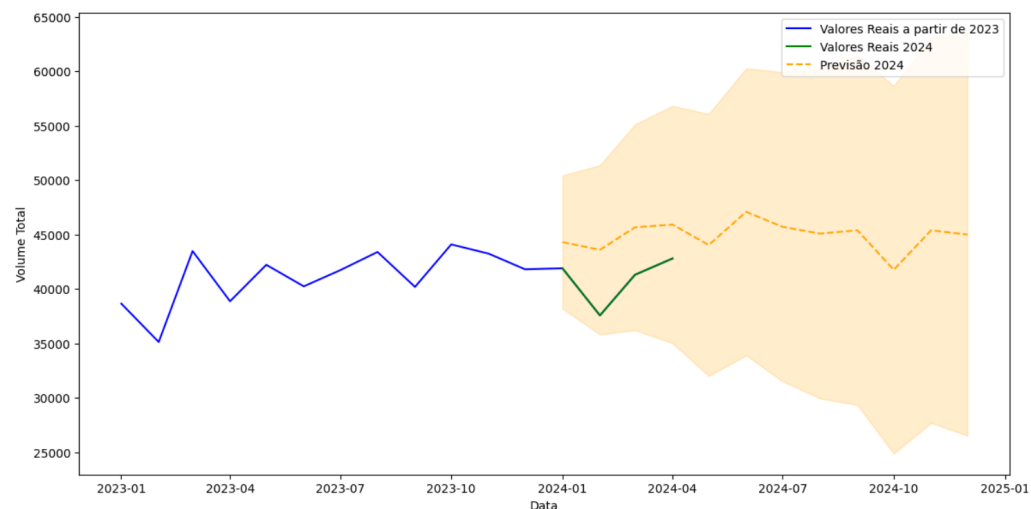


Figura 12. Modelo 2024. Fonte: próprio autor

5. Conclusão

Com base em nossos resultados, podemos concluir que os modelos usados, ARMA, ARIMA e SARIMA são todos capazes de gerar de forma eficaz uma previsão para o volume de tráfego em praças de pedágio. O modelo SARIMA, após a otimização dos hiperparâmetros, demonstrou ser o mais eficiente, com o menor erro de previsão. Isso se deve à capacidade do SARIMA de capturar padrões sazonais presentes nos dados de tráfego. É possível observar também que devido às características inerentes às séries temporais as previsões realizadas logo após o fim dos dados reais possuem uma maior confiabilidade. De forma que para a obtenção de dados mais precisos o ideal seria que o modelo fosse atualizado de forma mensal, garantido assim previsões assertivas.

Dado que o erro médio absoluto de nosso modelo é de aproximadamente 2100 veículos por mês acreditamos que os valores gerados podem possibilitar a empresa que gerencia o posto de pedágio uma estimativa valiosa para preparar o posto para a demanda esperada de modo a evitar problemas de gargalo no atendimento gerando transtornos diversos. É importante destacar também a importância do pré-processamento dos dados e da otimização dos hiperparâmetros para melhorar a precisão dos modelos preditivos. Além disso, a escolha adequada dos dados e a eliminação de outliers contribuíram para a qualidade dos resultados obtidos.

5.1. Trabalhos Futuros

Como trabalhos futuros, sugerimos a busca por bases de dados que possuam uma granularidade menor permitindo ajustes mais refinados à previsão do volume de tráfego. É possível também que sejam incluídas variáveis adicionais, como condições climáticas e eventos especiais, para enriquecer o modelo e potencialmente aumentar sua precisão. A implementação de um sistema de previsão em tempo real também pode ser explorada, fornecendo uma ferramenta valiosa para a gestão e planejamento do tráfego em rodovias.

Referências

- [Barth and Boriboonsomsin 2008] Barth, M. and Boriboonsomsin, K. (2008). Real-world carbon dioxide impacts of traffic congestion. *Transportation research record*, 2058(1):163–171.
- [Caleffi and Cybis] Caleffi, F. and Cybis, H. B. B. Aperfeiçoamento de modelo de previsão de conflitos em rodovias através de deep learning.
- [da Infraestrutura 2021] da Infraestrutura, M. (2021). Plano nacional de logística 2035. Acessado: 2024-06-04.
- [Gordin et al. 2011] Gordin, E., Klodzinski, J., and Santos, C. D. (2011). Safety benefits from deployment of open road tolling for main-line toll plazas in florida. *Transportation research record*, 2229(1):85–92.
- [Mendonça 2022] Mendonça, O. H. A. (2022). Radar: um mecanismo de predição de fluxo de tráfego de veículos em vias urbanas com base em aprendizado federado.
- [Monard and Baranauskas 2003] Monard, M. C. and Baranauskas, J. A. (2003). Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):32.
- [Neri and Silva Júnior 2023] Neri, T. B. and Silva Júnior, C. A. P. d. (2023). O efeito barreira de rodovia urbana e os impactos sobre a mobilidade e despesas com transportes. *urbe. Revista Brasileira de Gestão Urbana*, 15:e20210264.
- [Ramis and Santos 2012] Ramis, J. E. and Santos, E. A. d. (2012). Uso de automóveis e o caos urbano: considerações sobre o planejamento de transportes das grandes cidades. *Journal of Transport Literature*, 6:164–177.
- [SANTOS JÚNIOR et al. 2022] SANTOS JÚNIOR, D. S. d. O. et al. (2022). Método de ensemble para correção de modelos arima: uma abordagem de sistema híbrido para previsão de séries temporais.
- [SILVA] SILVA, L. d. Análise da aplicação do modelo arima: estudo em uma instituição federal de ensino superior. 2017.
- [Yang et al. 2012] Yang, H., Ozbay, K., and Bartın, B. (2012). Effects of open road tolling on safety performance of freeway mainline toll plazas. *Transportation research record*, 2324(1):101–109.
- [Zhao et al. 2019] Zhao, W., Gao, Y., Ji, T., Wan, X., Ye, F., and Bai, G. (2019). Deep temporal convolutional networks for short-term traffic flow forecasting. *Ieee Access*, 7:114496–114507.