



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
CURSO DE ENGENHARIA ELÉTRICA

Matheus Lenzi dos Santos

**Algoritmo de modelagem de perfis de condução com dados veiculares:
uma abordagem de aprendizado de máquina**

Florianópolis
2024

Matheus Lenzi dos Santos

**Algoritmo de modelagem de perfis de condução com dados veiculares:
uma abordagem de aprendizado de máquina**

Trabalho de Conclusão de Curso submetido ao curso de Engenharia Elétrica do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de bacharel em Engenharia Elétrica.

Orientador: Prof. Antonio Augusto Medeiros Fröhlich, Ph.D.

Florianópolis

2024

Santos, Matheus Lenzi dos

Algoritmo de modelagem de perfis de condução com dados veiculares: uma abordagem de aprendizado de máquina / Matheus Lenzi dos Santos ; orientador, Antonio Augusto Medeiros Fröhlich, 2024.

51 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia Elétrica, Florianópolis, 2024.

Inclui referências.

1. Engenharia Elétrica. 2. Aprendizado de máquina. 3. Conectividade veicular. 4. Simulação computacional. 5. Perfil de condução. I. Fröhlich, Antonio Augusto Medeiros. Universidade Federal de Santa Catarina. Graduação em Engenharia Elétrica. III. Título.

Matheus Lenzi dos Santos

**Algoritmo de modelagem de perfis de condução com dados veiculares:
uma abordagem de aprendizado de máquina**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de “Bacharel em Engenharia Elétrica” e aceito, em sua forma final, pelo Curso de Graduação em Engenharia Elétrica.

Florianópolis, 03 de julho de 2024.

Prof. Miguel Moreto, Dr.
Coordenador do Curso de Graduação em Engenharia Elétrica

Banca Examinadora:

Prof. Antonio Augusto Medeiros Fröhlich, Ph.D.
Orientador
Universidade Federal de Santa Catarina

Prof. Joceli Mayer, Ph.D.
Universidade Federal de Santa Catarina

Leonardo Passig Horstmann, MSc.
Laboratório de Integração de Software e Hardware

Este trabalho é dedicado aos meus queridos pais, familiares e amigos que me apoiaram durante toda a minha trajetória.

AGRADECIMENTOS

Agradeço a Deus e a todos que contribuíram para a minha jornada acadêmica e para o resultado deste trabalho.

Aos meus pais, Paulo Sergio e Juliana, expresso minha profunda gratidão pela excelente formação que recebi. Vocês são meus modelos, e seu amor incondicional e apoio incessante foram fundamentais para minha evolução e aprendizado.

Ao meu irmão, Pedro, que me ensinou muito e é sinônimo de dedicação, mostrando-me que não devemos ter medo de sonhar grande. Estendo meus agradecimentos a todos os meus familiares, especialmente aos meus avós e padrinhos, por serem a base de tudo, minha família.

À minha namorada, Julia, agradeço por me acompanhar em todos os momentos e por ser a melhor companhia que poderia ter tido nesta trajetória. Obrigado pelo apoio constante, pela compreensão e pelo amor sem igual. Você é minha maior inspiração e motivação.

Agradeço aos amigos pelo companheirismo ao longo desta jornada. Os momentos de descontração, trocas de conhecimento e apoio nos tornaram pessoas e profissionais melhores.

Aos professores e funcionários da universidade, agradeço não apenas pelo conteúdo e pelas condições oferecidas, mas também pelas orientações que me tornaram melhor. Obrigado por serem mestres e mentores, proporcionando-me a melhor formação.

Agradeço especialmente ao meu orientador, Prof. Guto, por todo o tempo dedicado a mim e ao meu trabalho. Sua sabedoria, dedicação e *feedbacks* foram de imensa importância para minha jornada. Você é uma referência inigualável, com um compromisso genuíno com a educação e a pesquisa.

Por fim, agradeço a todos que, de alguma forma, contribuíram para esta conquista. Este trabalho não seria possível sem o apoio e carinho de cada um de vocês. Muito obrigado!

RESUMO

Este trabalho tem como objetivo desenvolver um algoritmo de perfilamento de condutores utilizando técnicas de aprendizado de máquina baseadas em dados veiculares. Para isso, fez-se uso de uma simulação computacional para gerar dados representativos dos perfis de condução. A metodologia envolveu a criação de um ambiente controlado usando o software Simulation of Urban Mobility (SUMO), onde foram gerados cenários para a coleta de dados de condução. Estes dados foram então pré-processados e utilizados para treinar modelos de aprendizado de máquina para identificar padrões e agrupar perfis de condução. O processo de simulação resultou em um conjunto de dados com aproximadamente 100 milhões de pontos de dados, representando diversos cenários de condução. A escolha do simulador SUMO se mostrou vantajosa devido à sua capacidade de gerar cenários completos e oferecer um conjunto de dados ajustado para o desenvolvimento de algoritmos de inteligência artificial. Os modelos de aprendizado de máquina utilizados incluíram técnicas de agrupamento, como o K-means, e redes neurais recorrentes, como as Long Short-Term Memory (LSTM), para análise de séries temporais. Os resultados indicaram que a abordagem proposta é eficaz na modelagem de perfis de condutores, possibilitando a personalização de experiências de condução e contribuindo para a melhoria no ecossistema de mobilidade. Este trabalho estabelece uma base sólida que pode ser replicada e aprimorada para diversas aplicações específicas.

Palavras-chave: aprendizado de máquina, simulação computacional, conectividade veicular, perfil de condução.

ABSTRACT

This work aims to develop a driver profiling algorithm using machine learning techniques based on vehicular data. A computational simulation was employed to generate representative data for driving profiles. The methodology involved creating a controlled environment using the Simulation of Urban Mobility (SUMO) software, where scenarios were generated for driving data collection. This data was then preprocessed and used to train machine learning models to identify patterns and cluster driving profiles. The simulation process resulted in a dataset with approximately 100 million data points, representing diverse driving scenarios. The choice of the SUMO simulator proved advantageous due to its ability to generate comprehensive scenarios and provide a dataset tailored for the development of artificial intelligence algorithms. The machine learning models used included clustering techniques, such as K-means, and recurrent neural networks, like Long Short-Term Memory (LSTM), for time series analysis. The results indicated that the proposed approach is effective in modeling driver profiles, enabling personalized driving experiences and contributing to improvements in the mobility ecosystem. This work establishes a solid foundation that can be replicated and enhanced for various specific applications.

Key-words: machine learning, computational simulation, vehicular connectivity, driving profile.

LISTA DE FIGURAS

Figura 1 – Fluxograma da Metodologia	18
Figura 2 – Luxembourg SUMO Traffic (LuST)	21
Figura 3 – Cruzamento em Luxembourg SUMO Traffic (LuST).....	22
Figura 4 – Histograma do Tipo de Veículo	25
Figura 5 – Histograma da Duração das Viagens em Períodos de Simulação	26
Figura 6 – Boxplot da Soma da Duração das Viagens agrupada por Tipo Veículo ..	26
Figura 7 – Métricas para Avaliação do Número de Grupos	28
Figura 8 – Gráfico de Dispersão dos Grupos por Variáveis Escaladas	29
Figura 9 – Exemplo de LSTM de Passo Único	31
Figura 10 – Gráficos de Barras do Conjunto de Dados	33
Figura 11 – Configuração da LSTM com Propriedade de Portas	34

LISTA DE QUADROS

Quadro 1 – Estudos sobre Modelos de Perfilamento de Condução	14
Quadro 2 – Conjunto de Dados Resultante da Simulação <i>Emissions</i> do SUMO	19
Quadro 3 – Colunas do Conjunto de Dados da Simulação <i>Emissions</i> do SUMO	20
Quadro 4 – Arquivo para Algoritmo de Geração dos Tipos de Veículos	23
Quadro 5 – Arquivo Resultante do Algoritmo de Geração dos Tipos de Veículos.....	23
Quadro 6 – Arquivo Resultante do Algoritmo de Geração das Rotas.....	24
Quadro 7 – Algoritmo de Pré-processamento dos Dados	27
Quadro 8 – Médias dos Valores Numéricos Médios das Viagens por Grupo	30
Quadro 9 – Resultados da LSTM com Janela de Tamanho 10 para Classificação..	34

LISTA DE ABREVIATURAS E SIGLAS

AM Aprendizado de Máquina
C2X Comunicação de Veículos
C3X The McKinsey Connected Car Customer Experience
ECUs Unidades de Controle Eletrônicas
GPS Global Positioning System
IA Inteligência Artificial
IMU Inertial Measurement Unit
LiDAR Light Detection and Ranging
LSTM Long Short-Term Memory
LuST Luxembourg SUMO Traffic
OBD On Board Diagnostics
RADAR Radio Detection and Ranging
RNN Redes Neurais Recorrentes
SAE Sociedade de Engenheiros Automotivos
SUMO Simulation of Urban Mobility
V2I Veículo-a-Infraestrutura
V2V Veículo-a-Veículo

SUMÁRIO

1	INTRODUÇÃO.....	6
2	MOTIVAÇÃO	8
3	JUSTIFICATIVA	10
4	OBJETIVOS.....	12
5	REVISÃO DE LITERATURA	13
6	FUNDAMENTAÇÃO TEÓRICA.....	15
7	METODOLOGIA	17
8	DESENVOLVIMENTO	19
8.1	CONJUNTO DE DADOS	19
8.2	REDE DE MOBILIDADE VEICULAR	20
8.3	GERAÇÃO DOS TIPOS DE VEÍCULOS.....	22
8.4	GERAÇÃO DAS ROTAS DE VIAGEM.....	24
8.5	SIMULAÇÃO	24
8.6	ANÁLISE EXPLORATÓRIA DOS DADOS.....	25
8.7	PRÉ-PROCESSAMENTO DOS DADOS	27
8.8	AGRUPAMENTO DOS PERFIS DAS VIAGENS	27
8.9	PREDIÇÃO DO PERFIL EM SÉRIES TEMPORAIS	30
9	RESULTADOS E DISCUSSÕES.....	36
10	CONCLUSÃO	39
11	REFERÊNCIAS	40

1 INTRODUÇÃO

Um veículo conectado é um sistema de transporte interligado que permite a troca de dados entre automóveis, pedestres, infraestrutura e outros dispositivos móveis para melhorar a mobilidade. Alguns estudos mostram que os veículos conectados podem melhorar significativamente o fluxo de tráfego, reduzir os tempos de viagem e diminuir as emissões de gases de efeito estufa, além de proporcionar uma condução mais segura através da comunicação veículo-a-veículo (V2V) e veículo-a-infraestrutura (V2I) (OLIA et al., 2016).

Os veículos conectados de última geração, como o Lucid Air, têm a capacidade de produzir uma vasta quantidade de dados a partir de sensores como câmeras, *Light Detection and Ranging* (LiDAR), *Radio Detection and Ranging* (RADAR), *Global Positioning System* (GPS) e *Inertial Measurement Unit* (IMU). Esse volume massivo de sinais apresenta desafios significativos para as montadoras, que precisam enfrentar questões relacionadas ao armazenamento, processamento e transmissão dessas informações (DICKERT, 2023). Esses dados também representam uma oportunidade revolucionária para o ecossistema de mobilidade, permitindo melhorias substanciais nos serviços oferecidos aos proprietários de veículos, enquanto influenciam em diversos aspectos operacionais do veículo e seu entorno (BERTONCELLO et al., 2021).

Atualmente, os sistemas eletrônicos e digitais desempenham um papel central no funcionamento dos veículos, monitorando todos os principais sistemas, como motor, freios e diversos itens relacionados à segurança do condutor. Os automóveis possuem uma grande variedade de diferentes sensores e atuadores que possibilitam o correto funcionamento de sistemas sofisticados de segurança e assistência a condução, como aviso de colisão e prevenção de acidentes. Com a crescente complexidade dos veículos, as Unidades de Controle Eletrônicas (ECUs) se tornam um componente fundamental, pois gerenciam grandes quantidades de dados (KABIR, 2023). Quando devidamente processados, os dados gerados, armazenados e transmitidos proporcionam valiosos insights tanto sobre o veículo quanto sobre o motorista (GEORGAKOS et al., 2013).

Conscientes da importância dos dados gerados pelos veículos modernos, as montadoras adotaram sistemas que permitem o envio de dados em tempo real das ECUs para servidores na nuvem (KOWALIK; SZPYRKA, 2019). Esses sistemas utilizam várias tecnologias de rede para garantir a comunicação eficiente e segura dos

dados coletados, promovendo diagnósticos remotos e atualizações de software necessárias para manter a *performance* e a segurança dos veículos (SHANMUGAM, 2019).

A conectividade veicular está presente mesmo em veículos mais antigos. A utilização de *hardware after-market*, como os *dongles On Board Diagnostics* (OBD), possibilita a aquisição em tempo real dos dados do veículo. A porta OBD estabelecida pela Sociedade de Engenheiros Automotivos (SAE) na década de 1980, evoluindo para sua segunda versão na década de 1990, desempenhou um papel fundamental ao estabelecer um padrão para a transferência de dados dos sensores para fontes externas, tornando-se um componente presente em todos os novos veículos produzidos (SÜZEN; KAYAALP, 2018).

Em um âmbito global, a conectividade de dados veicular deixou de ser uma tendência e tornou-se realidade em vários países. Segundo a pesquisa *Connected Car Forecast 2021* realizada pela *SBD Automotive*, os mercados dos Estados Unidos, Europa e Japão estão experimentando um crescimento notável na adoção de veículos conectados. No mercado norte-americano, por exemplo, em 2022, aproximadamente 95% dos veículos novos eram nativamente conectados, ou seja, saíam de fábrica com a capacidade de enviar os dados gerados pelos sensores e controlados pelas centrais eletrônicas para a nuvem. A previsão é que a Europa alcance 100% de conectividade em sua nova frota ainda em 2024, enquanto o mercado japonês prevê que 90% dos novos veículos estarão conectados até 2025 (SBD AUTOMOTIVE, 2021).

2 MOTIVAÇÃO

Para os consumidores, a conectividade veicular assume uma importância cada vez maior. Segundo a pesquisa *Auto leaders prepare to seize big opportunities* da KPMG em 2022, mais de 30% dos executivos do setor automotivo consideram fatores como segurança de dados, recursos de infoentretenimento e conectividade veicular como de extrema importância em decisões de compra de veículos nos próximos cinco anos (SILBERG, 2022). Além disso, a consultoria *McKinsey & Company* destaca que os proprietários de veículos veem um alto potencial na conectividade, com até 56% deles, em alguns países, considerando a possibilidade de mudar de montadora para obter os benefícios dessa tecnologia (BERTONCELLO et al., 2021).

Nesse contexto de transformação, surgem tanto oportunidades quanto desafios para os participantes do ecossistema automotivo de dados. A *McKinsey & Company* aponta que a dificuldade de criar diferenciação nos produtos para atrair os consumidores e a falta de uma estrutura organizacional e ecossistema adequados para a monetização desses dados são algumas das principais barreiras a serem superadas para a adoção plena das funcionalidades dos veículos conectados (BERTONCELLO et al., 2021).

Com o objetivo de aprimorar a experiência dos motoristas e agregar valor a esses dados de ponta a ponta, as partes interessadas do setor automotivo estão desenvolvendo estratégias conjuntas para criar um fluxo de processamento de dados abrangente, que inclui a coleta, limpeza, padronização, distribuição, enriquecimento e outros tratamentos que possam beneficiar toda a cadeia. Uma dessas estratégias é o *framework* denominado *The McKinsey Connected Car Customer Experience (C3X)* que identifica cinco níveis de experiência do usuário, começando pelo monitoramento básico do veículo e chegando a um nível de inteligência em conexão com o ambiente externo (BERTONCELLO et al., 2021).

Seguindo a tendência tecnológica dos veículos conectados, fabricantes de automóveis estão buscando integrar ferramentas e técnicas de Inteligência Artificial (IA) em seus processos, visando extrair o máximo de informação dos dados disponíveis no setor. A previsão é que o mercado de IA na indústria automobilística deverá alcançar 215 bilhões de dólares até 2025, com uma taxa de crescimento anual estimada em 119% na adoção de ferramentas baseadas em Aprendizado de Máquina

(AM), aplicadas em diversas funções, como o monitoramento dos veículos e motoristas (BLAISDELL, 2019).

A aplicação de IA tem o potencial de melhorar a segurança, a eficiência e a confiabilidade de veículos conectados, o que possibilita o surgimento de novas modalidades de negócios, como o seguro sob demanda, em que as seguradoras podem precificar a apólice pelo modo de condução e uso do motorista. Além dos aspectos sociais de segurança e redução de acidentes, uma implementação de inteligência gera um diferencial competitivo, atraindo a atenção do consumidor. Na pesquisa realizada pela KPMG, 46% dos executivos do setor automotivo acreditam que as montadoras vão entrar com sucesso nesta área, vendendo os dados para as seguradoras (SILBERG, 2022).

A montadora de veículos *Tesla* já oferece como serviço junto a venda dos seus veículos, seguros com prêmio variável conforme a pontuação atribuída a condução do motorista. Segundo consta em seu site, o veículo é capaz de rastrear o comportamento de direção em tempo real e oferecer, diante desta avaliação, um seguro baseado no uso do proprietário (TESLA, 2023). Ao iniciar este serviço, *Tesla Insurance*, a montadora atribui uma pontuação inicial, que é usada para definir o prêmio. Conforme o veículo é dirigido, a pontuações de segurança diária é combinada em uma média ponderada de quilometragem de 30 dias para calcular a pontuação de segurança usada para atualizar o prêmio (TESLA, 2023).

O *Safety Score Beta* da *Tesla* é uma avaliação do comportamento ao dirigir com base em diversas métricas chamadas fatores de segurança, que são combinados para estimar a probabilidade de que a direção de determinado motorista possa resultar em uma colisão futura. Os fatores de segurança são: avisos de colisão frontal, frenagem brusca, curvas agressivas, seguimento inseguro do veículo da frente, excesso de velocidade, direção noturna, não-uso do cinto de segurança e a desativação forçada do piloto automático (TESLA, 2023).

3 JUSTIFICATIVA

O cenário de mobilidade brasileiro está carente de avanços na segurança do trânsito. De acordo com o Registro Nacional de Acidentes e Estatísticas de Trânsito, somente em 2022, ocorreram mais de um milhão de acidentes de trânsito no Brasil, resultando tragicamente em aproximadamente 21 mil mortes (MINISTÉRIO DOS TRANSPORTES, 2022). Um estudo de 2018 do Ministério dos Transportes, Portos e Aviação sobre Segurança nas Rodovias Federais revelou que 53,7% desses acidentes foram causados pelo chamado "fator humano". A negligência ou imprudência dos motoristas, seja por desrespeito às leis de trânsito (30,3%) ou falta de atenção (23,4%), são as principais causas de mortes nas rodovias federais, tornando a segurança no trânsito uma preocupação crescente (MINISTÉRIO DOS TRANSPORTES, 2018).

Portanto, é possível considerar a utilização dos dados gerados pelos veículos e seus condutores como uma fonte potencial de informação que possibilite a melhoria do ecossistema de mobilidade, com redução dos acidentes ou novas oportunidades de negócio. Além disso, a utilização dos dados dos veículos para a definição do perfil de condução do motorista permite, entre outras coisas, o desenvolvimento de experiências personalizadas para cada indivíduo. Para a construção de um sistema de inteligência consistente relacionado aos veículos conectados, é preciso desenvolver um processo escalável de processamento dos dados, contexto em que o presente trabalho se insere.

Para tanto, um conjunto de dados significativo de veículos é parte fundamental na construção de modelos de inteligência. A geração de dados reais de veículos com foco em perfilamento de condutores é uma tarefa extensa e custosa, que envolve múltiplas iterações de motoristas, veículos e cenários (ROMERA et al., 2016). A maioria dos bons conjuntos de dados sobre veículos não é pública. Aqueles que são públicos geralmente não foram produzidos por testes especificamente projetados para esse propósito, carecem de explicações detalhadas sobre as condições de contorno ou utilizam dados coletados por sensores de *smartphones* fixados no veículo, ou seja, não são dados veiculares (SALEH et al., 2017).

Logo, tendo o desafio de perfilar condutores, deve-se considerar a geração de dados a partir de simuladores de padrão industrial para atingir um conjunto de dados relevante para os modelos de IA, levando em conta aspectos como a reprodutibilidade dos cenários de simulação e a heterogeneidade do conjunto de

dados. A *Waymo*, empresa de veículos autônomos do *Google*, por exemplo, desenvolveu uma *SimulationCity*, que permite criar cenários realistas de simulação baseados em dados reais. Esse sistema é utilizado para testar e aprimorar a *performance* dos veículos autônomos em condições variadas, inclusive situações raras, sendo crucial para a construção de um conjunto de dados relevantes para o treinamento de modelos robustos de IA (WAYMO, 2021).

Além disso, a utilização de ML nesse contexto se deve ao fato de que, considerando o volume massivo de dados, com, por exemplo, centenas de diferentes variáveis de sensores, é preciso estabelecer uma ferramenta robusta capaz de processá-los e encontrar padrões além dos explícitos, que uma fórmula simples poderia não identificar. Ainda, com a previsão do aumento do número de veículos conectados e, por esse motivo, a maior disponibilidade de dados, a utilização de IA abre a possibilidade de evolução e adaptação dos modelos e compreende aspectos como a redução da dimensionalidade dos dados.

4 OBJETIVOS

O objetivo geral deste estudo é desenvolver um algoritmo para perfilamento de condutores utilizando inteligência artificial, especialmente aprendizado de máquina, com base em dados veiculares. Especificamente, busca-se:

- Avaliar a geração de dados veiculares por meio de simulador computacional.
- Analisar a qualidade e quantidade dos dados veiculares gerados.
- Customizar cenários de simulação para obter dados relevantes e representativos.
- Desenvolver e implementar algoritmos de pré-processamento de dados veiculares.
- Aplicar técnicas de aprendizado de máquina para o desenvolvimento de modelos de perfilamento de condutores.
- Validar os modelos desenvolvidos através de testes em cenários simulados.
- Propor estratégias para a implementação prática dos modelos e, quando possível, com dados reais.

Este estudo não pretende criar um algoritmo completamente novo, mas apresentar uma abordagem holística para o desenvolvimento de cenários de simulação, conjuntos de dados, e algoritmos de pré-processamento e aprendizado de máquina.

5 REVISÃO DE LITERATURA

As pesquisas recentes sobre esse tema têm se concentrado na aplicação de inteligência artificial para extrair informações relevantes desses dados e prever o comportamento de motoristas. Assim como pode ser visto no Quadro 1, existem diversos estudos sobre o desenvolvimento de modelos de perfilamento de condução, que variam em termos da aplicação prática, dos parâmetros analisados, das técnicas de predição utilizadas, das métricas de desempenho e do resultado esperado.

Em 2022, Zheng e colaboradores propuseram um sistema de classificação de estilos de direção em tempo real, baseado em observações de curto prazo. Durante a fase de treinamento, dados históricos são coletados e agrupados em categorias de estilos de direção por meio de aprendizado não supervisionado. Já na fase de inferência, os dados são transformados para o domínio da frequência antes do agrupamento, o que permite uma melhor compreensão do comportamento de direção. Além disso, a redução de dimensão é aplicada para simplificar o processo de agrupamento, resultando em uma classificação consistente dos dados do sensor de condução (ZHENG et al., 2022).

Em 2017, Saleh e colaboradores propuseram uma abordagem baseada em Redes Neurais Recorrentes (RNN) *Long Short-Term Memory* (LSTM) para classificação do comportamento de direção. Utilizando dados brutos de sensores, eles estruturam o problema como uma tarefa de classificação de séries temporais, classificando janelas de sequência de dados de sensores fundidos em diferentes classes de comportamento de direção. Essa metodologia permite uma análise detalhada do comportamento do condutor ao longo do tempo, facilitando a identificação de padrões de condução (SALEH et al., 2017).

Em 2019, Ping e colaboradores investigaram o impacto do comportamento do condutor no consumo de combustível, empregando algoritmos de *clustering* espectral não supervisionados para analisar a relação entre diferentes comportamentos de direção e o consumo de combustível. Eles destacam a importância de compreender essas relações para o desenvolvimento de sistemas avançados de assistência à condução. Além disso, propuseram um método de predição de curto prazo usando redes neurais LSTM supervisionadas, visando fornecer estratégias de condução eficientes em tempo real (PING et al., 2019).

Em 2019, Allison e Stanton contribuíram para o perfilamento de condução ao explorar o papel do *feedback* na redução de emissões através de comportamentos de

condução cotidianos. Ao focar a eficácia de cursos de direção e do *feedback* contínuo na sustentação de comportamentos eficientes, o estudo ressalta a importância de compreender e incentivar padrões de condução favoráveis para a redução de emissões, para a otimização do desempenho do veículo e a segurança do condutor (ALLISON; STANTON, 2019).

Em 2021, Peppes e colaboradores propuseram uma abordagem para análise de comportamento de condução utilizando métodos de aprendizado de máquina e aprendizado profundo para fluxos contínuos de dados veiculares. Eles desenvolveram uma plataforma integrada para coleta, armazenamento e análise de dados, aplicando técnicas de *clustering* para classificar o comportamento dos condutores e comparando diferentes algoritmos de aprendizado supervisionado. A integração de técnicas de extração de características e algoritmos de *clustering*, como o *K-means*, permitiu uma análise eficiente dos padrões de direção, e a avaliação de desempenho de métodos tradicionais de aprendizado de máquina e profundos destacou a eficácia dessas abordagens no perfilamento de condução (PEPPES et al., 2021).

Quadro 1 – Estudos sobre Modelos de Perfilamento de Condução

Referência	Metodologia	Objetivo	Resultados
Saleh et al., 2017	Classificação do comportamento de direção usando RNN LSTM	Classificar o comportamento de direção em séries temporais	Identificação detalhada de padrões de condução ao longo do tempo
Ping et al., 2019	Análise do impacto do comportamento do condutor no consumo de combustível utilizando algoritmos de <i>clustering</i> espectral e redes neurais LSTM	Analisar a relação entre comportamento de direção e consumo de combustível	Predição de estratégias de condução eficientes em tempo real
Allison; Stanton, 2019	Exploração do papel do feedback na redução de emissões através de cursos de direção e <i>feedback</i> contínuo	Reduzir emissões através de comportamentos de condução eficientes	Compreensão e incentivo de padrões de condução favoráveis
Peppes et al., 2021	Análise de comportamento de condução utilizando aprendizado de máquina e aprendizado profundo, com técnicas de <i>clustering</i>	Coletar, armazenar e analisar dados veiculares para classificar comportamento de condução	Comparação eficaz de algoritmos de aprendizado supervisionado e não supervisionado
Zheng et al., 2022	Classificação de estilos de direção em tempo real utilizando aprendizado não supervisionado, transformação de domínio da frequência e redução de dimensão	Melhorar a compreensão e a classificação de estilos de direção	Classificação consistente dos dados do sensor de condução

Fonte: o autor.

6 FUNDAMENTAÇÃO TEÓRICA

Dentre os trabalhos correlatos, o método de aprendizado não supervisionado mais comum é a análise de *cluster* (agrupamento), utilizada para análise exploratória a fim de encontrar padrões nos dados. O *clustering* agrupa os dados em subconjuntos de forma que objetos dentro do mesmo grupo (ou *cluster*) sejam mais similares entre si do que aqueles em grupos diferentes. O algoritmo *K-means* é uma técnica de agrupamento amplamente utilizada em aprendizado de máquina e análise de dados. Seu objetivo é dividir um conjunto de dados em K grupos distintos, onde cada ponto de dados pertence ao grupo cujo centroide (ponto médio do grupo) é o mais próximo. O processo envolve a inicialização de K centroides aleatórios, a atribuição de cada ponto de dados ao centroide mais próximo e a atualização dos centroides com base na média dos pontos atribuídos a eles. Este processo é repetido iterativamente até que os centroides se estabilizem e não mudem mais significativamente (JAMES et al., 2023).

Já em aprendizado supervisionado, as redes neurais são escolha comum e modelos computacionais inspirados na estrutura do cérebro humano, compostos por camadas de nós (neurônios) que processam dados de entrada e produzem saídas. Dentre as diversas arquiteturas de redes neurais, as RNNs são projetadas para processar sequências de dados, tornando-as ideais para tarefas de séries temporais. As redes neurais recorrentes podem ser subdivididas em duas tarefas: classificação e regressão. A classificação utiliza dados rotulados para aprender um mapeamento das entradas para rótulos de classe, visando aprender alguma função de decisão. E a regressão abrange métodos supervisionados que utilizam pares de entradas e saídas para aprender a prever saídas contínuas para novas entradas. As redes LSTM são uma variante das RNN projetadas para modelar sequências temporais e capturar dependências de longo prazo nos dados. Elas possuem uma arquitetura que inclui "portas" de entrada, saída e esquecimento. Essas portas regulam o fluxo de informações, permitindo que as LSTMs armazenem e acessem informações relevantes por períodos prolongados. As LSTMs são amplamente utilizadas em tarefas que envolvem dados sequenciais, como previsão de séries temporais (JAMES et al., 2023).

As séries temporais referem-se a uma sequência de dados pontuais ordenados no tempo, frequentemente utilizadas em análises de tendências e monitoramento de eventos ao longo do tempo. As redes LSTM são eficazes na

modelagem de séries temporais devido à sua capacidade de capturar dependências de longo prazo nos dados. Existem dois principais tipos de modelos LSTM aplicados em séries temporais: o modelo de passo único e o modelo de várias etapas. O modelo de passo único prevê o próximo valor na sequência com base nos valores anteriores, sendo ideal para previsões imediatas e curtas. Já o modelo de várias etapas prevê múltiplos pontos futuros em uma única execução, permitindo a criação de previsões mais abrangentes e de longo prazo (TENSORFLOW, 2022).

Considerando que esses são os algoritmos mais utilizados dentre os trabalhos correlatos para a tarefa de perfilamento e o objetivo de construir uma abordagem de ponta-a-ponta para esse desafio, sem necessariamente desenvolver modelos novos, essas foram as ferramentas de IA selecionadas para o desenvolvimento deste trabalho.

7 METODOLOGIA

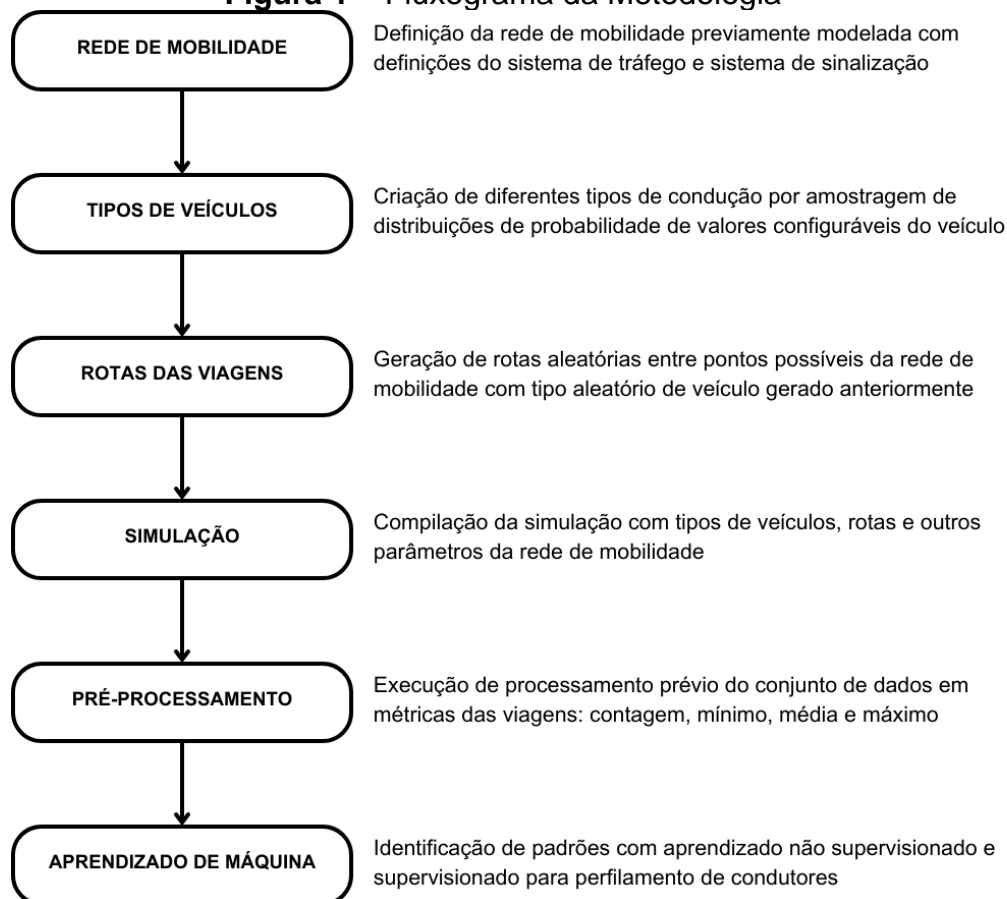
O primeiro passo para definição do perfil de condução dos motoristas foi criar um ambiente controlado de simulação para gerar conjuntos de dados que possam ser usados para treinar os algoritmos. Para isso, foi escolhido a ferramenta da simulação computacional *Simulation of Urban Mobility* (SUMO). Este software de código aberto, mantido pelo *openMobility Working Group*, composto por empresa como a *Bosch*, *AVL*, *Vector*, e institutos com o *Fraunhofer*, foi desenvolvido pelo Instituto de Sistemas de Transporte do Centro Aeroespacial Alemão. É um pacote de simulação de tráfego multimodal microscópico e contínuo, projetado para lidar com grandes redes de mobilidade urbana. Possui a capacidade de simular veículos autônomos, Comunicação de Veículos (C2X), tráfego multimodal, com carros, ônibus, trens, bicicletas, pedestres, transporte público, e geração de demanda e rotas (SUMO DOCUMENTATION, 2023).

A escolha desta ferramenta se deu pela facilidade de utilização aliada à capacidade de geração de diversos cenários complexos e realistas que produzem conjuntos de dados consistentes para o desenvolvimento de algoritmos. Esse simulador permite um grande número de customizações, com as quais é possível gerar diferentes perfis de condução do veículo (SUMO DOCUMENTATION, 2023). Adicionalmente à capacidade de personalização dos veículos, motoristas e cenários, o pacote de código aberto deste simulador oferece diversas ferramentas que permitem gerar diferentes configurações para a simulação, como rotas randomizada ou parâmetros de veículos que seguem uma distribuição de probabilidade (SUMO DOCUMENTATION, 2023). Deste modo, fazendo uso de tais ferramentas, é possível criar cenários reproduzíveis, que, com base no método científico, possam ser avaliados por outros pesquisadores.

A metodologia é apresentada de forma resumida na Figura 1, sendo que, dentre as ferramentas disponibilizadas pelo SUMO, foi utilizado um código em *Python* que permite a criação de diferentes tipos de condução por amostragem de distribuições de probabilidade de valores configuráveis do veículo. Deste modo, foram gerados perfis de condução não arbitrários para serem utilizados nos algoritmos de perfilamento. Os valores de média e desvio padrão para construção da distribuição normal de cada parâmetro seguiram o padrão de definido no simulador. Foi utilizado também uma ferramenta para geração de rotas de viagens aleatórias entre os pontos da rede de mobilidade escolhida. A combinação de rotas e perfis de condução não

determinísticos possibilitam a geração de um conjunto de dados heterogêneo, ideal para construção de algoritmos robustos (SUMO DOCUMENTATION, 2023).

Figura 1 – Fluxograma da Metodologia



Fonte: o autor.

8 DESENVOLVIMENTO

As ferramentas supracitadas foram utilizadas seguindo o processo descrito abaixo. Primeiro, foram gerados, a partir de distribuições normais de probabilidade dos parâmetros, tipos de veículos que representam diferentes estilos de condução. Para compor uma viagem, estes tipos de veículos foram selecionados aleatoriamente na geração de rotas, também aleatórias, entre os pontos da rede de mobilidade escolhida. Tendo definido as viagens, foi analisada a viabilidade da conclusão das mesmas, considerando as restrições do sistema de tráfego, como o sentido das vias. Por fim, as viagens validadas foram simuladas para produzir um conjunto de dados.

8.1 CONJUNTO DE DADOS

O simulador SUMO permite tipos de simulações que originam conjuntos de dados diferentes. Para esse trabalho, foi selecionado a simulação *Emissions*, que registra os valores dos “sensores” de todos os veículos para cada etapa da simulação e produz dados relevantes para o perfilamento de condução (SUMO DOCUMENTATION, 2023).

Quadro 2 – Conjunto de Dados Resultante da Simulação *Emissions* do SUMO

timestep	id	type	CO	CO2	HC	NOx	PMx	angle	eclass	electricity	fuel	lane	noise	pos	route	speed	waiting	x	y
0	0	89	165	2625	1	1	0	238	...	0	837	...	56	5	...	0	0	8559	8521
1	0	89	153	4179	1	2	0	238	...	0	1333	...	69	8	...	3	0	8557	8520
1	1	71	165	2625	1	1	0	315	...	0	837	...	56	5	...	0	0	4213	8175
2	0	89	124	3837	1	2	0	238	...	0	1224	...	64	12	...	4	0	8553	8518
2	1	71	148	3052	1	1	0	315	...	0	973	...	63	7	...	2	0	4212	8176
...

Fonte: o autor.

Presente, em parte, no Quadro 2, o conjunto de dados resultante da simulação possui 20 colunas que descrevem uma viagem. Entre os indexadores estão: *id*, identificador único do veículo que está realizando aquelas viagens, *type*, que faz referência ao tipo de veículo gerado e sorteado aleatoriamente para cumprir aquele trajeto, podendo ser repetido, visto que o número de viagens é superior ao número de perfis. E ainda, *timestep*, que indica o tempo da viagem na simulação. Por exemplo, se o veículo inicia a viagem no tempo 2 da simulação e termina no tempo

10, serão 9 registros de dados deste veículo específico. Como nem todos os veículos iniciam a viagem em momentos iguais e possuem rotas diferentes, as viagens tem duração variável, condicionadas também ao perfil de condução. Agrupando o indexador *id* e ordenando o *timestep* é possível chegar na descrição de todos os instantes da viagem de um veículo específico, com as séries temporais das variáveis categóricas e numéricas. No Quadro 3 estão todas as colunas da base de dados, uma breve descrição e a unidade de medida.

Quadro 3 – Colunas do Conjunto de Dados da Simulação *Emissions* do SUMO

Variável	Descrição	Unidade
timestep	Tempo da viagem na simulação	s
id	Identificador único do veículo (e da viagem)	ID
type	Identificador do tipo do veículo	ID
CO	Quantidade de monóxido de carbono emitido instantaneamente	mg/s
CO2	Quantidade de dióxido de carbono emitido instantaneamente	mg/s
HC	Quantidade de hidrocarbonetos emitido instantaneamente	mg/s
NOx	Quantidade de óxidos de nitrogênio emitido instantaneamente	mg/s
PMx	Quantidade de partículas em suspensão no ar instantaneamente	mg/s
angle	Ângulo do veículo em relação a referencia	°
eclass	Identificador da classe de emissão do veículo	ID
electricity	Quantidade de eletricidade consumida instantaneamente	Wh/s
fuel	Quantidade de combustível consumido instantaneamente	mg/s
lane	Identificador da faixa da rua	ID
noise	Nível de ruído emitido instantaneamente	dB
pos	Posição do veículo em relação a referencia	m
route	Identificador da rota do veículo	ID
speed	Velocidade instantanea	m/s
waiting	Tempo de espera parado	s
x	Coordenada x	-
y	Coordenada y	-

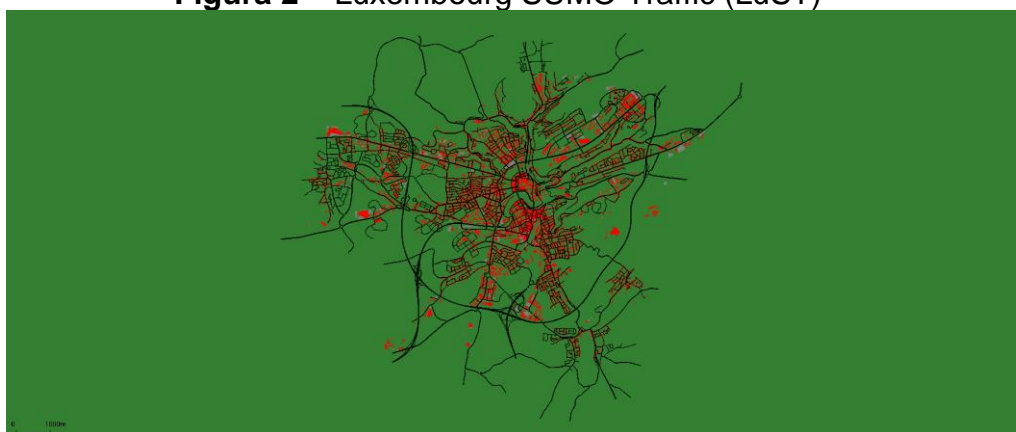
Fonte: o autor.

8.2 REDE DE MOBILIDADE VEICULAR

Como descrito na própria documentação da ferramenta, construir uma rede de mobilidade realista no SUMO é um processo trabalhoso. Envolve, entre outras coisas, considerar regras de trânsito local, como a velocidade máxima permitida, sentido das vias e número de faixas. Entretanto, existem excelentes cenários urbanos

disponíveis sob licença aberta na internet (SUMO DOCUMENTATION, 2023). Portanto, como rede de mobilidade para esse projeto, foi escolhido o cenário de simulação da região central de Luxemburgo, representado na Figura 2, que havia sido devidamente modelado no projeto de código aberto *Luxembourg SUMO Traffic* (LuST) (GITHUB, 2016). Este cenário possui, além da definição do sistema de tráfego (vias, sentido e número de faixas, etc.), todo o sistema de sinalização, com semáforos, conversões e outros elementos que propiciam a correta avaliação da demanda de tráfego, motivo pelo qual foi construído (CODECA et al., 2017).

Figura 2 – Luxembourg SUMO Traffic (LuST)



Fonte: o autor.

Além dos cenários distribuídos sob licença aberta, o SUMO permite a criação de um sistema de trânsito próprio, com a escolha arbitrária da malha viária, com ruas, interseções e outros elementos que compõem a rede (SUMO DOCUMENTATION, 2023). A escolha por utilizar o cenário previamente modelado de Luxemburgo ao invés de construir uma rede de mobilidade arbitrária teve como objetivo reduzir a subjetividade na definição da simulação de modo que, com o aumento da complexidade do sistema, a escolha dos parâmetros probabilísticos da criação do tipo de condução do veículo não fossem transportados diretamente para os resultados. Por exemplo, um tipo de veículo modelado com parâmetros de aceleração e velocidade mais elevados pode não poder atingir esse objetivo em determinados períodos da simulação devido ao trânsito em uma região mais congestionada da rede, como o cruzamento representado na Figura 3.

Figura 3 – Cruzamento em Luxembourg SUMO Traffic (LuST)

Fonte: o autor.

8.3 GERAÇÃO DOS TIPOS DE VEÍCULOS

Para a geração dos tipos de condução, foi necessário definir os atributos do veículo que seriam variados segundo uma distribuição de probabilidade normal, com parâmetros de média e desvio padrão escolhidos. Dentre os 37 parâmetros possíveis de serem alterados no SUMO, foram selecionados os 8 que influenciam no modelo de seguimento do veículo, como aceleração e desaceleração, e são mais relevantes para a definição do perfil de condução, excluindo aqueles que são puramente físicos, como comprimento e largura do veículo. Os parâmetros selecionados são: aceleração, desaceleração, velocidade máxima, multiplicadores, fator de desvio, impaciência e distância mínima em relação ao veículo à frente. Os valores dos parâmetros escolhidos seguiram os padrões pré-estabelecidos para o simulador, conforme mostrado no Quadro 4. É importante destacar que os parâmetros de classe do veículo e emissão (*vClass* e *emissionClass*) foram mantidos para todos os veículos, que eram carros de passeio a gasolina, preservando assim a base de comparação para o perfilamento de condutores (SUMO DOCUMENTATION, 2023).

Quadro 4 – Arquivo para Algoritmo de Geração dos Tipos de Veículos

<p>accel; normal(2.6,2.0) (m/ s²) aceleração técnica</p> <p>decel; normal(4.5,2.0) (m/ s²) desaceleração técnica</p> <p>maxSpeed; normal(55.55,27.62) (m/s) velocidade máxima técnica</p> <p>desiredMaxSpeed; normal(2778,1389) (m/s) velocidade máxima desejada</p> <p>speedFactor; normal(1.0,0.5) multiplicador do limite de velocidade da pista e desiredMaxSpeed</p> <p>speedDev; normal(0.5,0.2) [0.2,1.5] multiplicador do desvio do speedFactor</p> <p>impatience; normal(0.5,0.5) tendência dos motoristas para impedir veículos com maior prioridade</p> <p>minGap; normal(2.5,0.5) (m) distância mínima do veículo da frente</p>

Fonte: o autor.

O arquivo do Quadro 4 foi então processado por um algoritmo *Python* (*createVehTypeDistribution.py*) para gerar 100 diferentes tipos de veículos (*vType*) com valores segundo os limites estabelecidos pelas distribuições de probabilidade mencionadas (SUMO DOCUMENTATION, 2023). O Quadro 5 representa, em parte, o arquivo resultante desse processamento, que será utilizado na etapa de geração de rotas para selecionar os perfis para cumprir o trajeto na simulação.

Quadro 5 – Arquivo Resultante do Algoritmo de Geração dos Tipos de Veículos

```

<vTypeDistribution id="vehDist">

  <vType id="vehDist0" accel="2.919" decel="3.382" maxSpeed="90.207"
desiredMaxSpeed="2586.236" speedFactor="0.512" speedDev="0.641"
impatience="0.208" minGap="4.673"/>

  <vType id="vehDist1" accel="2.178" decel="7.160" maxSpeed="97.100"
desiredMaxSpeed="5485.851" speedFactor="2.156" speedDev="0.384"
impatience="0.870" minGap="4.335"/>

  <vType id="vehDist2" accel="3.078" decel="3.022" maxSpeed="29.598"
desiredMaxSpeed="1461.205" speedFactor="1.183" speedDev="0.430"
impatience="0.662" minGap="1.337"/>

  <vType id="vehDist3" accel="3.003" decel="3.527" maxSpeed="81.572"
desiredMaxSpeed="3390.855" speedFactor="1.657" speedDev="0.366"
impatience="0.536" minGap="4.275"/>

  <.../>

</vTypeDistribution>

```

Fonte: o autor.

8.4 GERAÇÃO DAS ROTAS DE VIAGEM

Para gerar o conjunto de viagens a serem realizadas na simulação, foi utilizado um algoritmo *Python (randomTrips.py)*, que escolhe pontos de origem e destino de maneira uniforme e aleatória, dada uma determinada rede de mobilidade. Nesse processo, o tipo de veículo (*vType*) foi incluído como parâmetro adicional para as viagens geradas (SUMO DOCUMENTATION, 2023). Dessa forma, o arquivo resultante, presente, em parte, no Quadro 6, descreve o momento e local de início da viagem, o ponto final, e o tipo de veículo dentre os gerados anteriormente. O tempo final de chegada da viagem depende de diversos fatores, como a rota e o perfil de condução, e só é conhecido ao final da simulação. Também foi utilizado um algoritmo *Python (duaiterate.py)* para eliminar problemas de conectividade e continuidade das rotas geradas, resultando em 7058 viagens validadas e aptas para serem simuladas (SUMO DOCUMENTATION, 2023).

Quadro 6 – Arquivo Resultante do Algoritmo de Geração das Rotas

```
<routes>

  <trip id="0" depart="0.00" from="-31070#7" to="--30402" type="vehDist"/>

  <trip id="1" depart="1.00" from="--31910#6" to="--31602#2" type="vehDist"/>

  <trip id="2" depart="2.00" from="-31622#6" to="-31346" type="vehDist"/>

  <trip id="3" depart="3.00" from="-32432#8" to="--30694#4" type="vehDist"/>

  <.../>

</routes>
```

Fonte: o autor.

8.5 SIMULAÇÃO

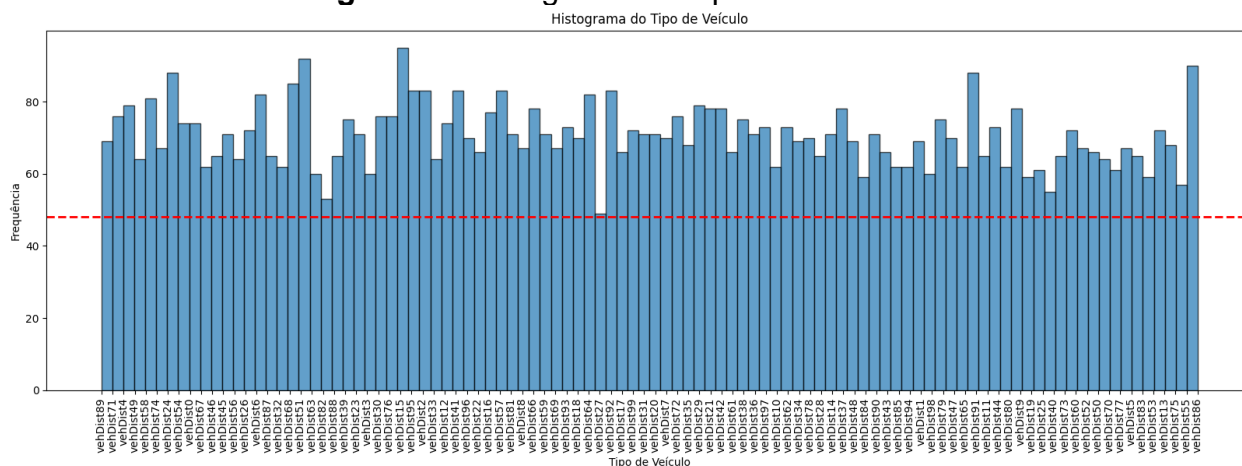
O processo de simulação foi então definido por um arquivo proprietário do SUMO, que estabelece os tipos de veículos e rotas explicados anteriormente, além de outros parâmetros que compõem a rede de mobilidade, como a temporização dos semáforos. A compilação da simulação durou aproximadamente 25 minutos (Processador: CPU @ 2.20GHz, 2208 MHz, 6 Núcleos, 12 Processadores Lógicos), e foram distribuídos 7058 veículos ao longo 81023 períodos de simulação (*timestep*). Adicionalmente, ocorreram 67 frenagens emergenciais e 6 colisões, que superam o

valor máximo de desaceleração pretendido pelo veículo. Esses eventos, frenagens emergenciais e colisões, não foram considerados para a construção dos modelos, mas poderiam ser explorados em trabalhos futuros.

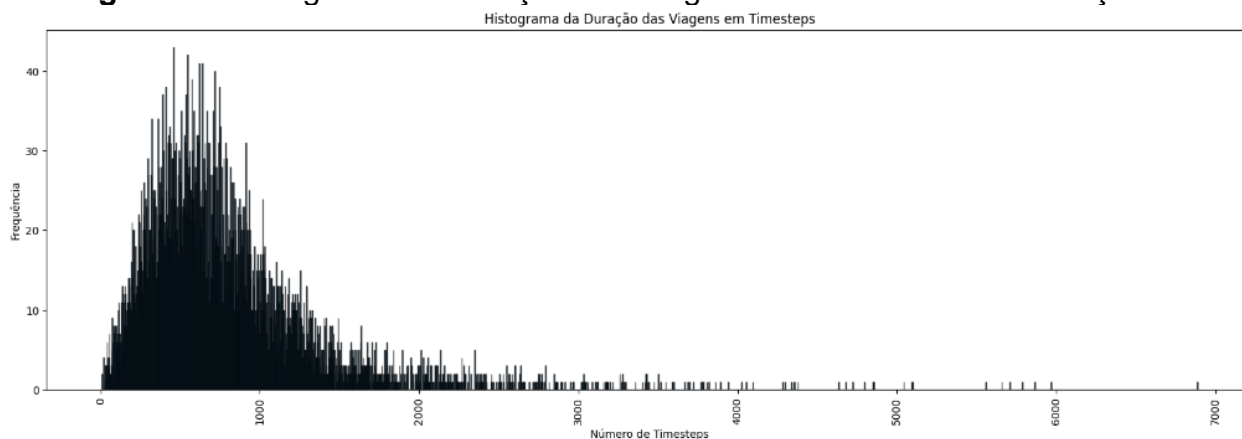
8.6 ANÁLISE EXPLORATÓRIA DOS DADOS

A simulação, conforme descrita acima, gerou um conjunto de dados com aproximadamente 100 milhões de pontos de dados, contendo os dados dos veículos durante as 7058 viagens realizadas. Cada veículo teve um tipo atribuído e foi instruído a cumprir a rota definida aleatoriamente. O histograma da Figura 4 apresenta a distribuição do tipo de veículo entre os que realizaram as viagens. É possível perceber que todos os perfis estão suficientemente representados e fizeram um número considerável de viagens para serem analisados.

Figura 4 – Histograma do Tipo de Veículo

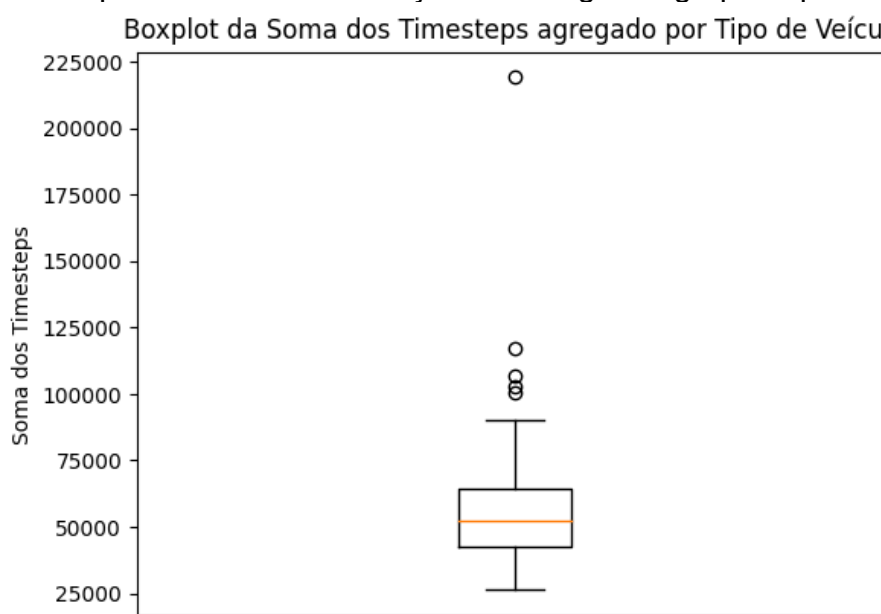


O histograma apresentado na Figura 5 representa a duração de todas as viagens realizadas, medida em períodos de simulação (*timestep*). É possível perceber uma distribuição normal com cauda longa, em que a duração da maioria das viagens fica concentrada ao redor da média, mas há casos em que a duração é muito maior. Esses valores distantes da média podem ocorrer devido ao perfil gerado, com uma velocidade menor, por exemplo, ou por algum caso específico da simulação. É importante reforçar que o início das viagens é distribuído durante toda a simulação, ou seja, nem todas as viagens começam e terminam no mesmo momento, o que ajuda a manter um fluxo de trânsito sem grandes engarrafamentos.

Figura 5 – Histograma da Duração das Viagens em Períodos de Simulação

Fonte: o autor.

No *boxplot* da Figura 6, é mostrado o resultado da soma do número de períodos de simulação (*timestep*) das viagens, agrupados por tipo de veículo gerado. É possível perceber que alguns tipos de veículos possuem durações de viagens que, quando somadas, resultam em uma maior representatividade em termos de tempo no conjunto de dados resultante da simulação. Essa conjuntura pode ser explicada pela natureza do perfil gerado, que pode ter acelerações menos bruscas ou uma velocidade menor, por exemplo.

Figura 6 – Boxplot da Soma da Duração das Viagens agrupada por Tipo Veículo

Fonte: o autor.

8.7 PRÉ-PROCESSAMENTO DOS DADOS

Considerando os 100 milhões de pontos de dados de todas as viagens da simulação, foi necessário estabelecer um processamento prévio dos dados para a construção dos modelos de aprendizado de máquina. Utilizando os indexadores disponíveis no conjunto de dados, os registros de cada instante de tempo foram agrupadas pelo identificador do veículo (*id*) e foram calculadas métricas como mínimos, máximos, médias e contagens das viagens. O algoritmo apresentado no Quadro 7 produziu um tabela resultante com 7058 linhas, ou seja, uma por cada viagem, e com 32 colunas, sendo 2 dessas identificadores (*id* e *type*) e as restantes a contagem dos passos de simulação (*timestep*), ou seja, a duração da viagem, além do mínimo, média e máximo das colunas numéricas do conjunto resultante da simulação.

Quadro 7 – Algoritmo de Pré-processamento dos Dados

```
metrics = trips.groupby("id").agg({
    "type": "first",
    "time": ["count"],
    "speed": ["min", "max", "mean"],
    "CO": ["min", "max", "mean"],
    "CO2": ["min", "max", "mean"],
    "HC": ["min", "max", "mean"],
    "NOx": ["min", "max", "mean"],
    "PMx": ["min", "max", "mean"],
    "electricity": ["min", "max", "mean"],
    "fuel": ["min", "max", "mean"],
    "noise": ["min", "max", "mean"],
    "waiting": ["min", "max", "mean"],
})
```

Fonte: o autor.

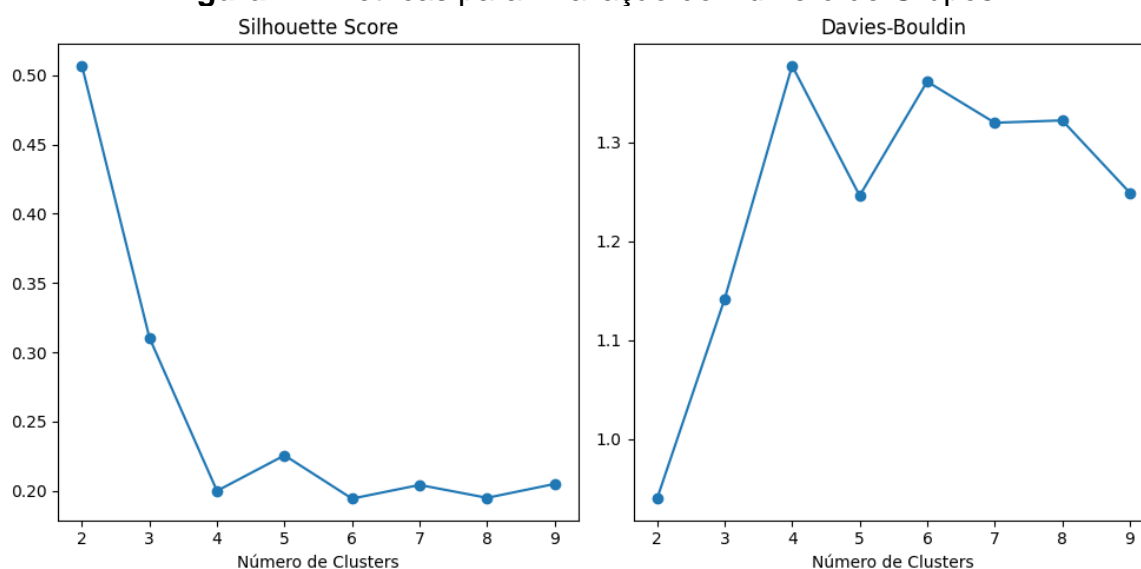
8.8 AGRUPAMENTO DOS PERFIS DAS VIAGENS

O conjunto de dados de métricas resultantes do pré-processamento da simulação foi submetido ao método de aprendizado não supervisionado mais comum, o agrupamento (*clustering*), em busca de identificar padrões que permitissem dividir as viagens segundo o perfil de condução. Seguindo a abordagem explorada em outros trabalhos apresentados na revisão de literatura, a ideia por trás da utilização da técnica *machine learning* não supervisionado baseia-se na possibilidade de, utilizando dados agregados de longo prazo, como métricas de viagens completas, encontrar

agrupamentos não explícitos nos dados através de um processo de aprendizado iterativo.

O algoritmo escolhido, *K-means*, possui uma implementação simplificada na biblioteca *scikit-learn* do *Python* e tem como principal parâmetro o número de grupos, *K*, que pode ser definido por um processo iterativo que avalia a qualidade do agrupamento dos dados (SCIKIT LEARN, 2024). As duas métricas utilizadas para essa avaliação foram os índices *Davies-Bouldin* e o *Silhouette Score* (SCIKIT LEARN, 2024). O índice *Davies-Bouldin* mede a compactação e separação dos *clusters*, onde valores menores indicam melhores agrupamentos. Já o *Silhouette Score* avalia a coesão interna e a separação entre clusters, com valores mais próximos de 1 indicando uma melhor estrutura de agrupamento.

Figura 7 – Métricas para Avaliação do Número de Grupos

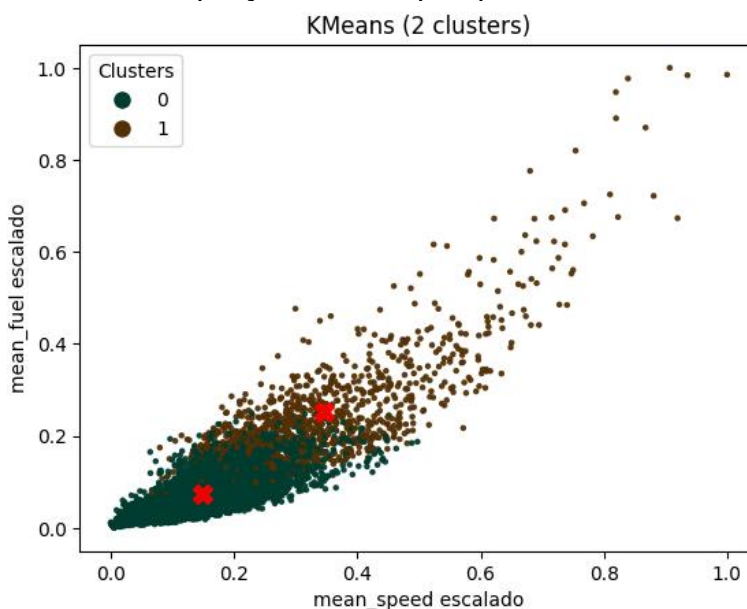


Fonte: o autor.

Dessa forma, como pode ser visto na Figura 7, o número de *clusters* que melhor divide os perfis nas viagens foi 2. É importante destacar que o conjunto de métricas foi escalado linearmente segundo o algoritmo *MinMaxScaler* para o treinamento do algoritmo (SCIKIT LEARN, 2024), com intuito de balancear a contribuição dos atributos, reescalando as colunas entre 0 e 1 com base nos valores mínimos e máximos de cada atributo. Na Figura 8, é possível conferir a divisão dos grupos para as variáveis escaladas de consumo de combustível instantâneo médio e velocidade instantânea média. Cada ponto representa uma viagem, as cores

correspondem aos dois diferentes agrupamentos, e os símbolos “X” representam os centros dos grupos.

Figura 8 – Gráfico de Dispersão dos Grupos por Variáveis Escaladas



Fonte: o autor.

Para entender o perfil de cada grupo, a Quadro 8 apresenta as médias dos valores numéricos médios das viagens. É possível perceber que o *cluster 0* possui muitas mais viagens do que o *cluster 1*. Em geral, as médias das variáveis numéricas no *cluster 1* são de 2 a 3 vezes maiores que no *cluster 0*. Pode-se inferir que o *cluster 1* contém os motoristas menos econômicos, mais poluentes ou mais agressivos, mas nenhum rótulo foi estabelecido, pois o objetivo do trabalho é criar uma abordagem holística para o desenvolvimento de cenários de simulação, conjuntos de dados, e algoritmos de pré-processamento e aprendizado de máquina. Logo, um desdobramento específico focado em ecologia, como descrito nos trabalhos da revisão de literatura, ou focado em segurança, como apresentado no exemplo da *Tesla*, seria possível a partir do processo estabelecido.

Quadro 8 – Médias dos Valores Numéricos Médios das Viagens por Grupo

cluster	Contagem	Média									
	id	mean_speed	mean_CO	mean_CO2	mean_HC	mean_NOx	mean_PMx	mean_electricity	mean_fuel	mean_noise	mean_waiting
0	6040	8	71	2.755	0,39	1	0,05	0	879	61	3
1	1018	17	139	5.926	0,78	2	0,14	0	1.890	67	2

Fonte: o autor.

8.9 PREDIÇÃO DO PERFIL EM SÉRIES TEMPORAIS

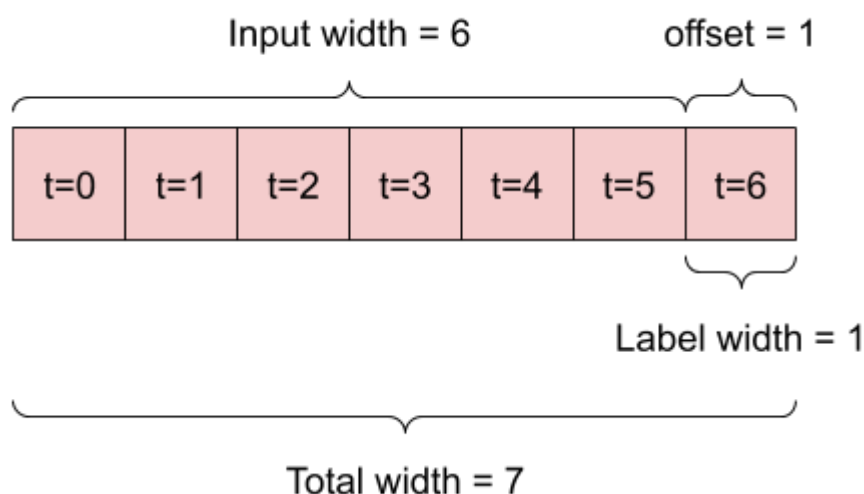
A etapa de processamento prévio dos dados em métricas das viagens produz uma redução significativa de informação de aproximadamente 5,6 milhões para 7 mil registros. Essa compressão das séries temporais em medidas de tendência (mínimo, máximo, média, contagem) possibilita a construção assertiva do modelo de *cluster*, mas deixa de aproveitar o potencial dos dados a cada instante das viagens. O agrupamento propicia um perfilamento através de aprendizado não supervisionado, utilizando esses dados agregados para encontrar padrões não explícitos nos dados. Já o aprendizado supervisionado necessita de dados rotulados para aprender um mapeamento das entradas para rótulos de classe. Então a etapa de *cluster* serve de insumo para a próxima etapa, fornecendo os rótulos para um treinamento supervisionado.

Assim como exposto na fundamentação teórica, as séries temporais possibilitam análises de tendências e monitoramento de eventos ao longo do tempo, e as LSTMs, como pode ser visto nos trabalhos citados nas referências bibliográficas e na própria fundamentação teórica, são boas em capturar essas dependências temporais dos dados (TENSORFLOW, 2022). Conectado com o objetivo do trabalho de propor estratégias para a implementação prática dos modelos, uma análise das séries temporais assim que os dados são gerados, sem ter que esperar a conclusão das viagens, tem diversas aplicações possíveis. No contexto de conectividade e comunicação veicular, os veículos compartilham entre si os dados que estão gerando. Assim, ser capaz de definir o perfil de condução dos veículos ao redor utilizando pequenas porções de séries temporais pode ajudar a ajustar a condução de veículos autônomos e não autônomos para aumentar a segurança, por exemplo.

Portanto, a ideia da aplicação de LSTMs no contexto deste trabalho, na sequência da etapa de *clustering*, tem como objetivo, dado uma janela de dados da série temporal de um veículo, indicar a qual dos *clusters* definidos na etapa anterior ele pertence. Novamente, os grupos 0 e 1 da etapa de agrupamento não foram rotulados segundo um critério específico, como segurança, pois o objetivo deste trabalho é desenvolver uma abordagem replicável e posteriormente customizável para cada cenário de aplicação.

Para a construção da LSTM, foram escolhidas como parâmetros de entrada (X) todas as variáveis numéricas do conjunto de dados das viagens, a saber: *CO*, *CO2*, *HC*, *NO*, *PMx*, *electricity*, *fuel*, *noise*, *pos*, *speed*, *waiting*, *x* e *y*. Relembrando, essas variáveis representam, em um dado tempo da viagem na simulação (*timestep*), o valor instantâneo gerado, emitido ou consumido por um veículo. Como variável alvo de saída do modelo (*y* ou *label*), foram escolhidos os *clusters*, que foram obtidos na etapa anterior e relacionados através do identificador único do veículo (*id*).

Figura 9 – Exemplo de LSTM de Passo Único



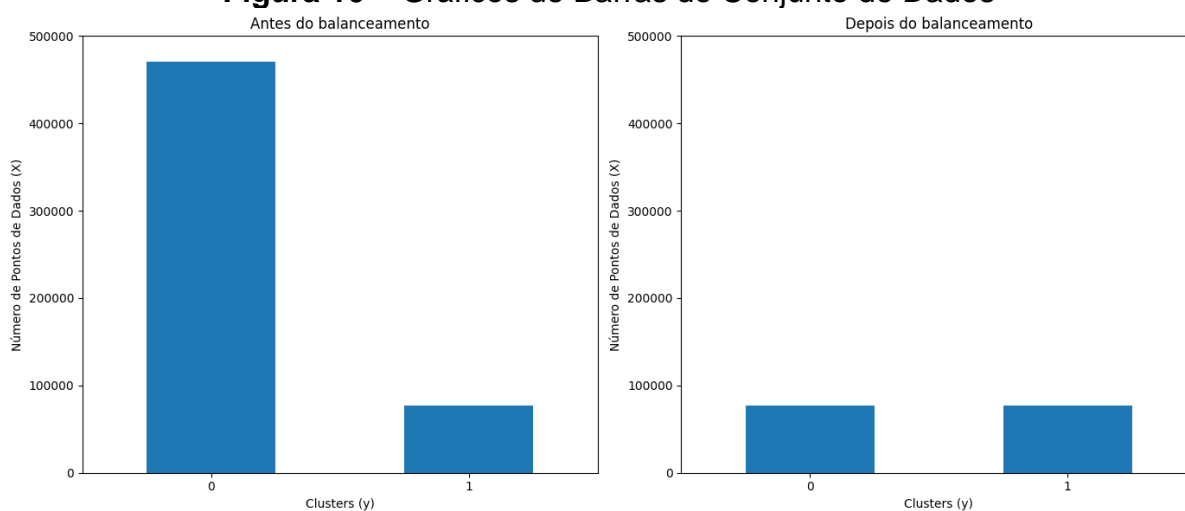
Fonte: TensorFlow (2022).

Dada a natureza do problema modelado, foi escolhida uma LSTM de passo único com uma janela de tamanho fixo, sem sobreposição. Assim como apresentado na Figura 9, a rede recebe um input de alguns períodos de tempo e pretende prever um rótulo (*label*) ao final do processamento. Ou seja, a LSTM recebe alguns registros, representando períodos de tempo, com colunas, variáveis numéricas descritas anteriormente, e prediz o *cluster* ao qual aquela série temporal pertence.

A escolha do tamanho da janela (*input width*) é um parâmetro importante na construção do conjunto de treinamento e do modelo, podendo ser definida tanto através de um processo iterativo que avalia o resultado do algoritmo para diferentes tamanhos quanto por um requisito da aplicação. A ideia de ter uma janela de tamanho fixo corrobora com o exemplo descrito anteriormente de prever o perfil de outro veículo a partir de uma porção de dados. Seria possível processar a série temporal inteira de uma viagem, mas isso não ajudaria a definir o perfil em tempo real, por exemplo.

Outro ponto importante na construção dos conjuntos de treino e teste é a prevenção contra o vazamento de informações entre esses conjuntos. Por exemplo, seguindo os preceitos de *machine learning*, não seria correto utilizar uma janela de uma série temporal de uma viagem no treinamento e outra janela da mesma série temporal da mesma viagem no teste. Isso resultaria na validação do modelo com exemplos semelhantes e já conhecidos.

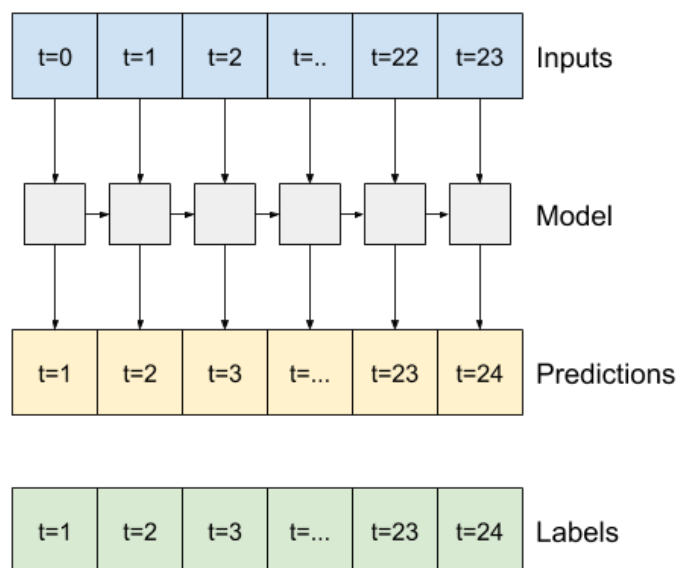
Dessa forma, a construção do *dataset* se deu da seguinte maneira: cada viagem foi individualmente dividida em janelas de 10 passos de simulação (*timestep*), e a cada uma dessas janelas foi atribuído o identificador único do veículo (*id*) e o grupo (*cluster*) ao qual pertencia. As janelas foram embaralhadas entre os conjuntos de treino e teste, mantendo a regra de que janelas de viagens do mesmo veículo só podem estar no mesmo conjunto. A atribuição do identificador do veículo serviu apenas para cumprir essa regra. Nesse caso, não era relevante avaliar a representatividade da quantidade de janelas por veículo, mas sim a quantidade de janelas por cluster. Por exemplo, os veículos com um perfil de velocidade média menor pertencentes ao *cluster* 0 podem ter viagens mais longas e, sendo o tamanho da janela um valor fixo, produzir mais dados de um só perfil. Além disso, considerando o desbalanceamento na divisão do agrupamento realizado na etapa anterior, onde o grupo 0 tinha 6 vezes mais representantes que o grupo 1, era preciso considerar o impacto disso no conjunto de dados desta próxima etapa. Dessa forma, a Figura 10 representa o balanceamento que foi feito no conjunto de dados para manter a equidade na representação dos dois grupos.

Figura 10 – Gráficos de Barras do Conjunto de Dados

Fonte: o autor.

O conjunto balanceado foi dividido em 80% para treinamento e 20% para testes, aproximadamente 120.000 e 30.000 amostras, respectivamente. As entradas (X) da rede eram tuplas com 120.000 amostras, cada uma contendo 10 registros para a janela de tamanho fixo e 13 colunas numéricas, no caso das entradas de treinamento.

Para a construção do modelo, foi considerado o sequenciamento de camadas LSTM. Portanto, foi utilizada a configuração apresentada na Figura 11, em que a rede faz previsões a cada entrada e utiliza suas propriedades de "portas" de entrada, saída e esquecimento para repassar essa informação para outras camadas. Como o objetivo do modelo era prever o *cluster* correspondente à série temporal de entrada, a última camada, que dá a saída, era totalmente conectada (*dense*), com ativação sigmoideal. Assim, o resultado produzido seria a classificação de 0 ou 1, representando os dois extremos correspondentes aos *clusters*.

Figura 11 – Configuração da LSTM com Propriedade de Portas

Fonte: TensorFlow, (2022).

É importante reforçar que o presente estudo não pretende criar um algoritmo completamente novo, portanto, não foi estabelecido nenhum *benchmarking* de resultado para o modelo. Dessa forma, uma *performance* melhor pode ser alcançada pela combinação de diferentes parâmetros do processo estabelecido até aqui. Por exemplo, a geração de mais viagens ou viagens mais longas, a definição de mais *clusters* no processo de aprendizado não supervisionado, uma rede mais robusta ou complexa no aprendizado supervisionado, ou uma janela de tamanho diferente aplicada ao conjunto de dados. No entanto, o objetivo deste trabalho era estabelecer uma abordagem holística de geração, processamento e treinamento para o perfilamento de condutores utilizando *machine learning*, que possa ser replicável e aplicável em diferentes cenários.

Quadro 9 – Resultados da LSTM com Janela de Tamanho 10 para Classificação

Métrica	Balanceado	Desbalanceado	Descrição
Precision	0,51	0,51	Proporção de verdadeiros positivos entre os positivos preditos. [1]
Recall	0,57	0,72	Proporção de verdadeiros positivos entre os casos reais positivos. [2]
F1-score	0,54	0,60	Média harmônica de precision e recall. [3]

[1], [2] e [3] TensorFlow, (2022)

Fonte: o autor.

A Quadro 9 apresenta os resultados da LSTM treinada com os conjuntos balanceado e desbalanceado, utilizando uma janela de tamanho fixo de 10. O caso desbalanceado apresentou um resultado superior, possivelmente por ter mais dados disponíveis para o treinamento e representar em maior quantidade o perfil mais abundante no conjunto de dados. Os resultados em ambos os casos foram inferiores ao esperado para um modelo de aprendizado de máquina, indicando que a rede poderia passar por um processo mais elaborado de construção e treinamento para atingir um resultado aceitável. Essa otimização do modelo pode servir como tema para um projeto futuro, onde, com uma abordagem abrangente já estabelecida, seria possível se concentrar na escolha dos parâmetros para obter um melhor resultado para uma aplicação específica. Uma outra questão relevante no conjunto de dados é a possibilidade de, dependendo da aplicação, fazer uma seleção dos atributos mais relevantes para a definição do perfil. Nesse caso, foram utilizadas as 13 variáveis numéricas disponíveis no *dataset*, mas a seleção criteriosa desses atributos pode melhorar a performance do modelo e sua aplicabilidade em diferentes cenários. Esta abordagem holística de geração, processamento e treinamento para o perfilamento de condutores utilizando *machine learning*, ainda que tenha espaço para otimizações, estabelece uma base sólida que pode ser replicada e aprimorada para diversas aplicações específicas.

9 RESULTADOS E DISCUSSÕES

Levando em consideração o objetivo estabelecido para este trabalho, é possível afirmar que foi alcançado um resultado consistente, caracterizado por um processo elaborado de geração de dados através de um simulador computacional, um algoritmo de pré-processamento dos dados e uma abordagem de *machine learning* composta por dois estágios. Esse processo, considerando todas as etapas, cumpre bem a tarefa de perfilamento de condutores.

Como mencionado anteriormente, cada uma das etapas poderia ser explorada mais profundamente para considerar todas as opções disponíveis de customização e melhoria. Esse detalhamento de cada etapa pode servir como tema para trabalhos futuros, onde um caso de uso específico, focado em uma aplicação particular, pode ser desenvolvido. Entretanto, este trabalho estabelece uma abordagem holística e replicável para o perfilamento de condutores usando *machine learning*, que pode ser adaptada e melhorada conforme necessário.

Especificamente sobre a geração de dados veiculares por meio de simulador computacional: a escolha do simulador SUMO se mostrou muito benéfica pela disponibilidade de ferramentas que propiciaram a geração de um cenário de simulação realista em um conjunto de dados heterogêneo e representativo para a tarefa de perfilamento. Como citado na metodologia, existem aspectos do simulador, como as colisões ou frenagens de emergência, que poderiam ser mais bem explorados para algum estudo específico, como de segurança, por exemplo. A qualidade e quantidade dos dados veiculares gerados ficaram bem evidenciadas pelos 100 milhões de pontos de dados, sendo 13 principais colunas numéricas. O conjunto de dados *Emissions* do SUMO possui muitas variáveis que seriam relevantes para casos de estudo ecológicos, por exemplo.

Fica como sugestão para trabalhos futuros o cruzamento das variáveis disponíveis em conjuntos de dados públicos de veículos reais e dos veículos gerados pela simulação computacional. Uma grande área de convergência reforçaria a relevância do processo desenvolvido por este trabalho. A avaliação baseada em veículos reais ajudaria a guiar a seleção das variáveis mais relevantes (e disponíveis) para os modelos de aprendizado de máquina. É importante ressaltar que a falta de padronização dos dados entre diferentes modelos e marcas de veículos é o principal desafio tecnológico para o uso dos sinais gerados pelos veículos conectados, seguido da necessidade de ter as ferramentas e habilidades para o uso efetivo e estratégico

dos dados (BERTONCELLO et al., 2023). Portanto, não existe uma única referência para os veículos reais, e os conjuntos de dados podem ter mais ou menos convergência.

Quanto à customização dos cenários de simulação, tanto a escolha entre as diversas redes de mobilidade disponíveis sob licença de código aberto quanto a escolha dos parâmetros utilizados para gerar os perfis podem alterar os resultados. Para esta iteração, foram escolhidos apenas 8 dos 37 parâmetros possíveis para geração dos perfis, mas entre os que não foram utilizados, vários podem ser relevantes para outras aplicações, abrindo margem para maiores customizações. Por exemplo, os parâmetros de classe do veículo e emissão podem ser úteis em projetos voltados à ecologia e à mobilidade elétrica.

Conectando com uma possibilidade de aplicação prática da geração de perfis, seria possível observar uma frota real, extrair e substituir os parâmetros da distribuição normal dos perfis, que neste trabalho seguiram o padrão do simulador, dessa frota real e preparar cenários de simulação baseados, considerando a rede de mobilidade, na frota real com perfis que representem esses motoristas conhecidos, gerando assim uma aumento e refinamento dos dados para posterior treinamento dos algoritmos de IA.

Sobre os algoritmos de pré-processamento de dados veiculares, é evidente a compressão na quantidade de dados de mais de 5 milhões para 7 mil registros. Como destacado no texto, esse processo teve vantagens e desvantagens. Entre as vantagens, propiciou um conjunto de dados suficientemente representativo para o treinamento eficiente de um modelo de aprendizado não supervisionado. Como desvantagem, perdeu informações relevantes sobre séries temporais que são efetivos em caracterizar tendências, apesar de terem sido recuperadas posteriormente com o aprendizado supervisionado. O pré-processamento dos dados se faz necessário quando, em uma simulação que pode ser considerada pequena, aproximadamente 7 mil viagens em 80 mil períodos de simulação, compilando por algo em torno de 20 minutos, resultou em mais de 5 milhões de registros. Em qualquer cenário mais complexo, onde mais perfis sejam considerados e mais viagens sejam simuladas, a quantidade de dados resultante da simulação precisará de um processamento prévio para tornar possível a sua utilização.

Quanto à aplicação de técnicas de aprendizado de máquina para o desenvolvimento de modelos de perfilamento de condutores, é possível afirmar que o

resultado foi satisfatório em termos de processo, mas poderia ser melhor explorado em termos de *performance*. O objetivo do trabalho foi concluído quando o encadeamento das etapas, desde a geração até a predição, ficou bem definido, claro e lógico. Fica evidente que a sequência de um modelo de aprendizado não supervisionado com dados de métricas, resultando em um agrupamento, e um modelo de aprendizado supervisionado com séries temporais, prevendo o rótulo encontrado anteriormente, poderiam ser posteriormente replicados para outros dados, até mesmo reais dos veículos. Seria possível ter um conjunto de dados já rotulados que permitisse um aprendizado supervisionado logo de primeira, mas o primeiro estágio com *K-means* permite identificar padrões não explícitos nos dados, que, se definidos arbitrariamente, poderiam não ser considerados. O agrupamento poderia ser mais bem elaborado se as variáveis de entrada fossem ponderadas pela relevância que devem ter na classificação de um perfil específico.

Quanto ao aprendizado supervisionado utilizando LSTMs, o problema de classificação dos rótulos encontrados no agrupamento a partir de janelas de tamanho fixo e sem sobreposição explicitou bem o processo para uma predição baseada em séries temporais. A rede e o modelo utilizados nesse cenário de aprendizado supervisionado serão melhor definidos quando, diante de uma aplicação real, tiverem requisitos funcionais e não-funcionais claros. Por exemplo, essa técnica poderia ser embarcada em um veículo e, portanto, deveria ser capaz de fazer predições em tempo de execução baixo com *hardware* limitado. Ou ainda, poderia estar rodando na nuvem, com alto poder computacional, mas para uma quantidade massiva de dados. Logo, em termos de *performance*, principalmente as LSTMs, podem ser melhor trabalhadas, mas tendo um caso de estudo, o desenvolvimento deve ter mais insumos para atingir o resultado esperado. No mais, os modelos foram validados em termos de entradas e saídas com os dados da simulação. É um tema interessante para próximos trabalhos avaliar o resultado dos modelos para outros cenários e configurações, por exemplo, uma outra rede de mobilidade disponível em código aberto.

10 CONCLUSÃO

Por fim, o presente trabalho cumpriu seu objetivo de estabelecer uma abordagem holística para o desenvolvimento de cenários de simulação, conjuntos de dados e algoritmos de pré-processamento e aprendizado de máquina. A pesquisa criou bases sólidas para que outros estudos possam utilizá-la como referência e promover avanços nesta área. A sequência de etapas, desde a geração até a predição, demonstrou que o encadeamento de modelos é uma abordagem eficaz e replicável.

Os modelos desenvolvidos foram validados com dados da simulação, evidenciando sua aplicabilidade. No entanto, a performance das LSTMs pode ser melhorada para atender diferentes requisitos funcionais e não-funcionais. A integração de dados reais de veículos para validar e aprimorar os modelos desenvolvidos é uma sugestão importante para trabalhos futuros.

O trabalho mostrou-se promissor e oferece um caminho claro para a aplicação prática das técnicas desenvolvidas, promovendo uma condução mais segura e personalizada. Avaliar o resultado dos modelos em outros cenários e configurações, bem como explorar redes de mobilidade disponíveis em código aberto, são caminhos interessantes para pesquisas futuras.

11 REFERÊNCIAS

ALLISON, Craig K; STANTON, Neville A. Eco-driving: the role of feedback in reducing emissions from everyday driving behaviours. *Theoretical Issues in Ergonomics Science*, Taylor & Francis, v. 20, n. 2, p. 85–104, 2019.

BERTONCELLO, Michele et al. Corporate business building to unlock value in automotive connectivity. McKinsey & Company, 2023. Disponível em: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/corporate-business-building-to-unlock-value-in-automotive-connectivity>. Acesso em: 24 nov. 2023.

BERTONCELLO, Michele et al. Unlocking the full life-cycle value from connected-car data. McKinsey & Company, 2021. Disponível em: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/unlocking-the-full-life-cycle-value-from-connected-car-data>. Acesso em: 23 nov. 2023.

BLAISDELL, Rick. Intelligent Cars: AI and the automotive industry. Rick's Cloud, 2019. Disponível em: <https://rickscloud.com/intelligent-cars-ai-and-the-automotive-industry/>. Acesso em: Acesso em: 18 nov. 2023.

CODECÁ, Lara et al. Luxembourg SUMO Traffic (LuST) Scenario: Traffic Demand Evaluation. *IEEE Intelligent Transportation Systems Magazine*, IEEE, v. 9, n. 2, p. 52–63, 2017

CONNECTED Car Forecast. SBD Automotive, 2021. Disponível em: <https://www.sbdautomotive.com/reports/connected-car-forecast>. Acesso em: Acesso em: 10 nov. 2023.

DICKERT, Chris. Network Overload? Adding Up the Data Produced by Connected Cars. *Visual Capitalist*, 2023. Disponível em: <https://www.visualcapitalist.com/network-overload/>. Acesso em: 20 nov. 2023.

GEORGAKOS, Georg et al. Reliability challenges for electric vehicles: From devices to architecture and systems software. In: PROCEEDINGS of the 50th Annual Design Automation Conference. [S.l.: s.n.], 2013. p. 1–9.

GITHUB. Icodeca/LuSTScenario: Luxembourg SUMO Traffic (LuST) Scenario, 2016. Disponível em: <https://github.com/icodeca/LuSTScenario>. Acesso em: 05 dez. 2023.

JAMES, Gareth et al. An introduction to statistical learning: With applications in python. [S.l.]: Springer Nature, 2023.

KABIR, Md Rafiul; RAY, Sandip. Virtualization for Automotive Safety and Security Exploration. In: 2023 IEEE 16th Dallas Circuits and Systems Conference (DCAS). [S.l.: s.n.], 2023. p. 1–4.

KOWALIK, Bartosz; SZPYRKA, Marcin. Online environment for data acquisition from car sensors. *Automatyka/Automatics*, v. 23, n. 1, p. 7–7, 2019.

MINISTÉRIO DOS TRANSPORTES. Avaliação da segurança viária nas rodovias federais, 2018. Disponível em: <https://www.gov.br/transportes/pt-br/assuntos/noticias/artigos/avaliacao-da-seguranca-viaria-nas-rodovias-federais>. Acesso em: 18 nov. 2023.

MINISTÉRIO DOS TRANSPORTES. Registro Nacional de Acidentes e Estatísticas de Trânsito, 2022. Disponível em: <https://www.gov.br/transportes/pt-br/assuntos/transito/conteudo-Senatran/registro-nacional-de-acidentes-e-estatisticas-de-transito>. Acesso em: 19 nov. 2023.

OLIA, Arash et al. Assessing the potential impacts of connected vehicles: mobility, environmental, and safety perspectives. *Journal of Intelligent Transportation Systems*, Taylor & Francis, v. 20, n. 3, p. 229–243, 2016.

PEPPES, Nikolaos et al. Driving behaviour analysis using machine and deep learning methods for continuous streams of vehicular data. *Sensors*, MDPI, v. 21, n. 14, p. 4704, 2021.

PING, Peng et al. Impact of driver behavior on fuel consumption: Classification, evaluation and prediction using machine learning. IEEE access, IEEE, v. 7, p. 78515–78532, 2019.

ROMERA, Eduardo et al. Need data for driver behaviour analysis? Presenting the public UAH-DriveSet. In: IEEE. 2016 IEEE 19th international conference on intelligent transportation systems (ITSC). [S.l.: s.n.], 2016. p. 387–392.

SALEH, Khaled; HOSSNY, Mohammed; NAHAVANDI, Saeid. Driving behavior classification based on sensor data fusion using LSTM recurrent neural networks. In: IEEE. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). [S.l.: s.n.], 2017. p. 1–6.

SCIKIT Learn. Davies Bouldin Score, 2024. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html. Acesso em: 07 fev. 2024.

SCIKIT Learn. KMeans, 2024. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans>. Acesso em: 05 fev. 2024.

SCIKIT Learn. MinMaxScaler, 2024. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>. Acesso em: 06 fev. 2024.

SCIKIT Learn. Silhouette Score, 2024. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html. Acesso em: 07 fev. 2024.

SHANMUGAM, Karthik. Securing inter-processor communication in automotive ECUs. [S.l.], 2019.

SILBERG, Gary. Auto leaders prepare to seize big opportunities. KPMG, 2022. Disponível em: <https://experience.kpmg.com/gaes-2022#/subpage/home>. Acesso em: 26 nov. 2023.

SUMO Documentation. Network Building, 2023. Disponível em: https://sumo.dlr.de/docs/index.html#network_building. Acesso em: 03 dez. 2023.

SUMO Documentation. Scenarios, 2023. Disponível em: <https://sumo.dlr.de/docs/Data/Scenarios.html>. Acesso em: 03 dez. 2023.

SUMO Documentation. Misc, 2023. Disponível em: <https://sumo.dlr.de/docs/Tools/Misc.html>. Acesso em: 06 dez. 2023.

SÜZEN, Ahmet Ali; KAYAALP, Kıyas. Web based tracking of vehicle fault and performance data on OBD II. Scientific Journal of Mehmet Akif Ersoy University, Burdur Mehmet Akif Ersoy University, v. 1, n. 1, p. 13–16, 2018.

TENSORFLOW. F1Score, 2024. Disponível em: https://www.tensorflow.org/addons/api_docs/python/tfa/metrics/F1Score. Acesso em: 16 jun. 2024.

TENSORFLOW. Precision, 2024. Disponível em: https://www.tensorflow.org/api_docs/python/tf/keras/metrics/Precision. Acesso em: 17 jun. 2024.

TENSORFLOW. Recall, 2024. Disponível em: https://www.tensorflow.org/api_docs/python/tf/keras/metrics/Recall. Acesso em: 17 jun. 2024.

TENSORFLOW. Time series forecasting, 2024. Disponível em: https://www.tensorflow.org/tutorials/structured_data/time_series. Acesso em: 17 jun. 2024.

TESLA. Insurance, 2023. Disponível em: <https://www.tesla.com/insurance>. Acesso em: 26 nov. 2023.

TESLA. Safety Score Beta, 2023. Disponível em: <https://www.tesla.com/support/safety-score#version-2.0>. Acesso em: 26 nov. 2023.

WAYMO. Simulation City: Introducing Waymo's most advanced simulation system yet for autonomous driving, 2021. Disponível em: <https://waymo.com/blog/2021/07/simulation-city/>. Acesso em: 28 nov. 2023.

ZHENG, Xihu et al. Real-time driving style classification based on short-term observations. IET Communications, Wiley Online Library, v. 16, n. 12, p. 1393–1402, 2022.