



DISSERTAÇÃO DE MESTRADO PROFISSIONAL

**Arquitetura Entrelaçada para
Predição de Próximo Sensor Externo em
Trajetória de Veículos Roubados**

GUSTAVO VIEIRA ISOBE DE MACEDO

Orientador Prof. Dr. Vinícius Pereira Gonçalves

Programa de Pós-Graduação Profissional em Engenharia Elétrica

DEPARTAMENTO DE ENGENHARIA ELÉTRICA
FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

**Arquitetura Entrelaçada para Predição de Próximo Sensor Externo em Trajetória de Veículos
Roubados**

Entangled Architecture for Predicting the Next External Sensor in Stolen Vehicle Trajectories

Gustavo Vieira Isobe de Macedo

Orientador: Prof. Dr. Vinícius Pereira Gonçalves, ENE/UnB

PUBLICAÇÃO: PPEE.MP.080

BRASÍLIA-DF

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO PROFISSIONAL

**Arquitetura Entrelaçada para
Predição de Próximo Sensor Externo em
Trajetória de Veículos Roubados**

GUSTAVO VIEIRA ISOBE DE MACEDO

Orientador Prof. Dr. Vinícius Pereira Gonçalves

*Dissertação de Mestrado Profissional submetida ao Departamento de Engenharia
Elétrica como requisito parcial para obtenção
do grau de Mestre em Engenharia Elétrica*

Banca Examinadora

Prof. Dr. Vinícius Pereira Gonçalves, FT/UnB
Presidente

Prof. Dr. João José Costa Gondim, FT/UnB
Examinador Interno

Prof. Dr. José Florêncio de Queiroz Neto, UFC
Examinador externo

Prof. Dr. Rafael Rabelo Nunes, FT/UnB
Suplente

FICHA CATALOGRÁFICA

MACEDO, GUSTAVO VIEIRA ISOBE

Arquitetura Entrelaçada para Predição de Próximo Sensor Externo em Trajetória de Veículos Roubados [Distrito Federal] 2025.

xvi, 69 p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2025).

Dissertação de Mestrado Profissional - Universidade de Brasília, Faculdade de Tecnologia.

Departamento de Engenharia Elétrica

- | | |
|--------------------------|---------------------------------|
| 1. Trajetória Veicular | 2. Trajetória de Sensor Externo |
| 3. Aprendizagem Profunda | 4. Arquitetura Entrelaçada |
| I. ENE/FT/UnB | II. Título (série) |

REFERÊNCIA BIBLIOGRÁFICA

MACEDO, G. V. I. (2025). *Arquitetura Entrelaçada para Predição de Próximo Sensor Externo em Trajetória de Veículos Roubados*. Dissertação de Mestrado Profissional, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 69 p.

CESSÃO DE DIREITOS

AUTOR: GUSTAVO VIEIRA ISOBE DE MACEDO

TÍTULO: Arquitetura Entrelaçada para Predição de Próximo Sensor Externo em Trajetória de Veículos Roubados .

GRAU: Mestre em Engenharia Elétrica ANO: 2025

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Do mesmo modo, a Universidade de Brasília tem permissão para divulgar este documento em biblioteca virtual, em formato que permita o acesso via redes de comunicação e a reprodução de cópias, desde que protegida a integridade do conteúdo dessas cópias e proibido o acesso a partes isoladas desse conteúdo. O autor reserva outros direitos de publicação e nenhuma parte deste documento pode ser reproduzida sem a autorização por escrito do autor.

GUSTAVO VIEIRA ISOBE DE MACEDO

Depto. de Engenharia Elétrica (ENE) - FT

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

CEP 70919-970 - Brasília - DF - Brasil

DEDICATÓRIA

Dedico esta jornada e conquista aos meus pais, irmãos, familiares e amigos, que sempre me incentivaram e apoiaram.

Dedico, de coração, esta jornada e conquista à minha esposa, parceira, amiga, confidente e amor da minha vida, Marcella, por todo o apoio, paciência, incentivo e compreensão diante do tempo investido neste Mestrado.

Dedico, também, ao meu filho Miguel, que, mesmo ainda sem entender, me proporcionou, ao longo desta jornada acadêmica, outro "Mestrado", em paralelo a este, ensinando-me sobre como cuidar de uma vida além da minha.

“Quem o acompanha na jornada é mais importante do que a chegada ao destino.”

Gustavo Vieira Isobe de Macedo

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Dr. Vinícius Pereira Gonçalves, pelos ensinamentos e pela condução talentosa e essencial para a conclusão desta jornada com sucesso.

Agradeço ao nosso grupo de orientação, que proporcionou ideias e contribuições essenciais para a qualidade dos conteúdos produzidos.

Agradeço aos colegas e amigos do Ministério da Justiça e Segurança Pública, que plantaram a semente do desejo de realizar um mestrado. Em especial, além das chefias Rodrigo Lange e Gustavo Henrique, e das equipes, ao PM de SP Fernandes, com quem tive a satisfação de trabalhar lado a lado e debater ideias sobre meu projeto de mestrado; ao perito da PF Peixoto, que revisou meu primeiro pré-projeto e me ensinou muito sobre o que é e como é a trajetória de um mestrado; e ao Murilo Goes de Almeida, que forneceu os dados anonimizados e auxiliou no entendimento de seus significados para a realização dos estudos.

Por fim, agradeço a Deus por iluminar a mim e a todos os envolvidos nesta trajetória.

RESUMO

Raramente um veículo roubado possui um equipamento de georreferenciamento acoplado para transmitir sua localização em tempo real e auxiliar na sua recuperação. No entanto, utilizando radares de velocidade e sistemas de reconhecimento óptico de caracteres (*Optical Character Recognition* - OCR), é possível identificar pontos discretos da trajetória desse veículo. Ao registrar esse tipo de sequência histórica, surge a necessidade de prever o próximo sensor externo da trajetória. Esse desafio vai além da predição em si, abrangendo também a definição da arquitetura de predição mais adequada. Parte deste trabalho segue a linha principal de outros estudos e fornece uma análise dos dados que influenciam o desempenho da predição, demonstrando que informações sobre qual via da rua o veículo foi identificado auxiliam na previsão do próximo sensor externo. Por outro lado, ao observar a ausência de um consenso acerca da arquitetura ideal para esse contexto, adota-se uma perspectiva mais ampla para identificar os principais critérios que influenciam a escolha da arquitetura mais adequada. Este estudo avalia a transição da arquitetura ideal em função do comprimento da sequência histórica e do formato da representação geográfica. Os resultados revelam uma transição na arquitetura de melhor desempenho ao variar o tamanho da sequência histórica, sendo o ponto de transição influenciado pelo tipo de representação geográfica das localizações. Como principal contribuição, este trabalho apresenta uma nova arquitetura entrelaçada de *deep learning*, denominada Transformer e LSTM Entrelaçados (TLE). Essa arquitetura TLE se sobressai em cenários de predição com curtas sequências históricas em relação a outras arquiteturas em seu estado-da-arte. Esses resultados são ainda mais vantajosos considerando o contexto de veículos roubados, em que o tempo é um fator crucial para sua recuperação.

Palavras-chave: Trajetória Veicular; Trajetória de Sensor Externo; Aprendizagem Profunda; Arquitetura Entrelaçada.

ABSTRACT

Stolen vehicles rarely have georeferencing devices to transmit real-time locations for recovery assistance. However, by using speed radars and Optical Character Recognition (OCR), it is possible to identify discrete points along the vehicle's trajectory. Recording such a historical sequence brings the need to predict the next external sensor in the trajectory, a challenge that extends beyond the prediction itself to determining the most suitable prediction architecture. On one hand, this work aligns with existing studies and provides data analyses that influence prediction performance, demonstrating that information about the street lane where the vehicle was identified aids in predicting the next external sensor. On the other hand, observing the lack of consensus on the ideal architecture for this context, a broader perspective is adopted to identify the key criteria influencing the choice of the most suitable architecture. This study evaluates the transition in optimal architecture as a function of historical sequence length and geographic representation format. The results reveal a shift in the best-performing architecture as the historical sequence length varies, with the shift point being influenced by the format of geographic representation of the locations. As a key contribution, this work introduces a novel entangled deep learning architecture, referred to here as the Transformer and LSTM Entangled (TLE). This TLE architecture outperforms other state-of-the-art architectures in prediction scenarios with short historical sequences. These findings offer significant advantages in the context of stolen vehicle geolocation, where time is a critical factor for successful recovery.

Keywords: Vehicle Trajectory; External Sensor Trajectory; Deep Learning; Entangled Architecture.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	OBJETIVOS	5
1.2	CONTRIBUIÇÕES DO TRABALHO	6
1.3	PUBLICAÇÕES	8
1.4	ESTRUTURA DA DISSERTAÇÃO	9
2	CONCEITOS.....	10
2.1	DEFINIÇÃO DO PROBLEMA.....	10
2.2	REFERENCIAL TEÓRICO	12
2.2.1	ARQUITETURAS DE APRENDIZADO DE MÁQUINA	12
2.2.2	REPRESENTAÇÃO GEOGRÁFICA.....	16
2.3	CONSIDERAÇÕES FINAIS	19
3	TRABALHOS RELACIONADOS	20
3.1	INCLUSÃO DE NOVAS INFORMAÇÕES PARA A PREDIÇÃO	20
3.2	NOVA ARQUITETURA DE PREDIÇÃO	23
3.3	CONSIDERAÇÕES FINAIS	28
4	METODOLOGIA E ARQUITETURA ENTRELAÇADA	29
4.1	FLUXOGRAMA DA METODOLOGIA	29
4.1.1	DEFINIÇÃO DO TRABALHO	29
4.1.2	DEFINIÇÃO DO ESCOPO.....	30
4.1.3	MENSURAÇÃO DOS EXPERIMENTOS	32
4.1.4	ANÁLISE DOS RESULTADOS	33
4.2	CONSIDERAÇÕES FINAIS	34
5	ARQUITETURA ENTRELAÇADA TLE.....	35
5.1	CONSIDERAÇÕES FINAIS	39
6	PRIMEIRO ESTUDO - GANHO DE PREDIÇÃO	40
6.1	INTRODUÇÃO	40
6.2	PRÉ-PROCESSAMENTO	40
6.3	ARQUITETURAS COMPARADAS - 1º ESTUDO.....	42
6.3.1	ARQUITETURA LSTM	43
6.3.2	ARQUITETURA TRANSFORMER	44
6.3.3	ARQUITETURA TLE.....	44
6.4	EXPERIMENTOS E ANÁLISES - 1º ESTUDO.....	45
6.5	CONSIDERAÇÕES FINAIS	47

7	SEGUNDO ESTUDO - APOIO À DEFINIÇÃO DE MELHOR ARQUITETURA	49
7.1	INTRODUÇÃO	49
7.2	PRÉ-PROCESSAMENTO	50
7.3	ARQUITETURAS COMPARADAS - 2º ESTUDO.....	52
7.3.1	ARQUITETURA XGBOOST	52
7.3.2	ARQUITETURA LSTM	53
7.3.3	ARQUITETURA TLE.....	53
7.4	EXPERIMENTOS E ANÁLISES - 2º ESTUDO.....	53
7.5	CONSIDERAÇÕES FINAIS	58
8	CONCLUSÃO	61
8.1	DISCUSSÕES FINAIS	61
8.2	LIMITAÇÕES	62
8.3	TRABALHOS FUTUROS	63
	REFERÊNCIAS BIBLIOGRÁFICAS	66

LISTA DE FIGURAS

1.1	Veículos roubados em São Paulo, Brasil.	2
1.2	Radares de velocidade mapeados em São Paulo, Brasil.	3
1.3	Sensores do CE.	4
1.4	Sensores de GO.	4
1.5	Sensores do DF.	4
2.1	Ilustração da predição do próximo sensor externo.	11
2.2	Exemplo de uma trajetória com 9 registros de sensores externos em SP.	12
2.3	Relação entre subcategorias de inteligência artificial e arquiteturas deste trabalho.	13
2.4	Funcionamento da arquitetura XGBOOST. Fonte: CHEN et al. (1).	13
2.5	Funcionamento da arquitetura LSTM. Fonte: TECH (2).	14
2.6	Funcionamento da arquitetura Transformer. Fonte: VASWANI et al. (3).	15
2.7	Exemplo da representação geográfica de um ponto P utilizando EPSG:4326 (WGS84). Fonte: MATHIAS et al. (4).	17
2.8	Exemplo de projeção cilíndrica em um mapa plano utilizando EPSG:3857 (Web Mercator). Fonte: GURGEL (5).	17
2.9	Exemplo de recorte de zonas e hemisférios utilizando EPSG:32723 (UTM). Fonte: GIO- VANINI (6).	18
4.1	Fluxograma da Metodologia.	29
5.1	Funcionamento da arquitetura TLE implementada.	37
6.1	Diagrama dos treinamentos.	42
6.2	Arquitetura LSTM implementada.	43
6.3	Arquitetura Transformer implementada.	44
6.4	Comparação entre as abordagens 1 (com ID sensor) e 2 (sem ID sensor) em relação ao tamanho da sequência histórica de entrada.	46
6.5	Comparação geral entre as abordagens 1 e 2.	46
6.6	Comparação entre as arquiteturas LSTM, TLE e Transformer em relação ao tamanho da sequência histórica de entrada.	47
6.7	Comparação geral entre as arquiteturas LSTM, TLE e Transformer.	47
7.1	Arquitetura XGBoost implementada.	52
7.2	Comparação das arquiteturas XGBoost, LSTM e TLE utilizando a métrica MSE.	55
7.3	Comparação das arquiteturas XGBoost, LSTM e TLE utilizando a métrica R^2	56
7.4	Comparação do formato de representação geográfica Web Mercator sobre WGS84 utili- zando as métricas MSE e R^2	58
7.5	Comparação do formato de representação geográfica UTM sobre WGS84 utilizando as métricas MSE e R^2	59

7.6	Comparação do formato de representação geográfica UTM sobre Web Mercator utilizando as métricas MSE e R^2	60
7.7	Ponto de transição das arquiteturas de melhor desempenho deslocado pela mudança de formato de representação geográfica.....	60
8.1	Distribuição dos registros de latitudes no Ceará (CE).	64
8.2	Distribuição dos registros de latitudes em Goiás (GO).	64
8.3	Distribuição dos registros de latitudes no Distrito Federal (DF).	64
8.4	Distribuição dos registros de latitudes em São Paulo (SP).....	64
8.5	Distribuição dos registros de longitudes no Ceará (CE).....	65
8.6	Distribuição dos registros de longitudes em Goiás (GO).	65
8.7	Distribuição dos registros de longitudes no Distrito Federal (DF).....	65
8.8	Distribuição dos registros de longitudes em São Paulo (SP).	65

LISTA DE TABELAS

3.1	Resumo comparativo entre estudos relacionados.	25
6.1	Quantidades de Predições.	43
6.2	Resultados das avaliações das predições das bases de teste CE, GO e DF.....	45
7.1	Comparação dos valores de MSE e R^2 para diferentes comprimentos de sequência nas arquiteturas XGBoost, LSTM e TLE.	54
7.2	Comparação das mudanças nos formatos de representação geográfica.....	57
7.3	Impacto das mudanças relativas nos formatos de representação geográfica ao longo dos comprimentos de sequência.	57

LISTA DE ABREVIACES

2D	<i>Duas Dimenses</i>
AdamW	<i>Adaptive Moment estimation with decoupled Weight decay</i>
AI	<i>Artificial Intelligence</i>
Attn-BiLSTM	<i>Attention Bidirectional Long Short-Term Memory</i>
Attn-LSTM	<i>Attention Long Short-Term Memory</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BERT MLM	<i>Bidirectional Encoder Representations from Transformers Masked Language Model</i>
BiLSTM	<i>Bidirectional Long Short-Term Memory</i>
BO	<i>Boletim de Ocorrncia</i>
CE	<i>Cear</i>
CICCN	<i>Centro Integrado de Comando e Controle Nacional</i>
CSBC	<i>Congresso da Sociedade Brasileira de Computao</i>
DBSCAN	<i>Density-Based Spatial Clustering of Applications with Noise</i>
DF	<i>Distrito Federal</i>
DINT	<i>Diretoria de Inteligncia</i>
EPSG	<i>European Petroleum Survey Group</i>
EST	<i>External Sensor Trajectory</i>
GB	<i>Gigabyte</i>
GBRT	<i>Gradient Boosted Regression Trees</i>
GELU	<i>Gaussian Error Linear Units</i>
GO	<i>Gois</i>
GPS	<i>Global Positioning System</i>
GPT-3	<i>Generative Pre-trained Transformer 3</i>

GRU	<i>Gated Recurrent Unit</i>
HighD	<i>Highway Drone dataset</i>
ID	<i>Identificador</i>
iNATran	<i>intention-aware Non-Autoregressive Transformer com aprendizado multi-attention</i>
IoV	<i>Internet of Vehicles</i>
JISA	<i>Journal of Internet Services and Applications</i>
KNN	<i>K-Nearest Neighbor</i>
LBERT	<i>LSTM using BERT MLM as an embedding layer</i>
LBERT-FT	<i>LSTM using BERT MLM as an embedding layer Fine-Tuned</i>
LDGST-LSTM	<i>Local Dynamic Graph Spatiotemporal–Long Short-Term Memory</i>
LSTM	<i>Long Short-Term Memory</i>
LW2V	<i>LSTM using Word2Vec as an embedding layer</i>
LW2V-FT	<i>LSTM using Word2Vec as an embedding layer Fine-Tuned</i>
MAE	<i>Mean Absolute Error</i>
MHA-LSTM	<i>Multi-Head Attention Long Short-Term Memory</i>
MJSP	<i>Ministério da Justiça e Segurança Pública</i>
MLP	<i>Multi-Layer Perceptron</i>
MNIST	<i>Modified National Institute of Standards and Technology dataset</i>
MSE	<i>Mean Squared Error</i>
MT	<i>Multi-Task</i>
MT-FI	<i>Multi-Task Full Imputation</i>
MT-FIC	<i>Multi-Task Full Imputation with Clustering</i>
MVE	<i>Maximum Velocity Error</i>
N	<i>Norte</i>
NGSIM	<i>Next Generation Simulation</i>

NIC	<i>Next Value Imputation with Clustering</i>
OCR	<i>Optical Character Recognition</i>
ODS	<i>Objetivo de Desenvolvimento Sustentável</i>
ONU	<i>Organização das Nações Unidas</i>
PLN	<i>Processamento de Linguagem Natural</i>
pNEUMA	<i>Large-scale dataset of naturalistic trajectories</i>
POI	<i>Point of Interest</i>
PRF	<i>Polícia Rodoviária Federal</i>
QP	<i>Questões de Pesquisa</i>
R^2	<i>Coeficiente de Determinação</i>
RAM	<i>Random Access Memory</i>
ReLU	<i>Rectified Linear Unit</i>
RF	<i>Random Forest</i>
RMSE	<i>Root Mean Squared Error</i>
RNN	<i>Recurrent Neural Network</i>
RQ	<i>Research Question</i>
S	<i>Sul</i>
SBC	<i>Sociedade Brasileira de Computação</i>
SBCUP	<i>Simpósio Brasileiro de Computação Ubíqua e Pervasiva</i>
SimpleRNN	<i>Simple Recurrent Neural Network</i>
SP	<i>São Paulo</i>
SSD	<i>Solid-State Drive</i>
TB	<i>Terabyte</i>
TLE	<i>Transformer and LSTM Entangled</i>
UTM	<i>Universal Transverse Mercator</i>

WGS84	<i>World Geodetic System 1984</i>
Word2Vec	<i>Word to Vector</i>
XGBoost	<i>Extreme Gradient Boosting</i>

1 INTRODUÇÃO

O Brasil, por suas proporções continentais e altos índices de criminalidade, tem sofrido com grande quantidade de roubo e de furto de veículos. De acordo com o 18º Anuário Brasileiro de Segurança Pública, publicado em 2024 (7), só no ano de 2023 foram 354.472 roubos e furtos de veículos em todo o território nacional, com uma taxa de 297,5 crimes dessas categorias a cada 100 mil veículos, chegando a uma média de cerca de 971 veículos retirados de seus legítimos proprietários todos os dias no Brasil em 2023. Assim, considerando um valor médio do veículo usado em cerca de 80 mil reais (8), em 2023 houve um prejuízo nacional para os cidadãos da ordem de 28 bilhões de reais. Ainda, em grande parte, os veículos roubados ou furtados são utilizados para a prática de outros crimes, como sequestro, roubo a bancos, transporte de drogas, armas ilegais, assassinatos, entre outros.

O Ministério da Justiça e Segurança Pública – MJSP é o órgão público máximo do Poder Executivo Federal responsável por políticas públicas em âmbito nacional que busquem o aprimoramento da segurança pública, tendo internamente departamentos voltados para esse tipo de problema, como a Polícia Rodoviária Federal – PRF.

Além do fato de que o patrulhamento das rodovias federais serem de responsabilidade da PRF, quando algum veículo atravessa a fronteira de um estado para outro, a Polícia Estadual de um estado não tem mais jurisdição para atuar no combate ao crime em outro estado. Para enfrentar esta realidade, a PRF tem criado programas de combate ao roubo e ao furto de veículos como o Sistema Nacional de Alarmes (9) e o Alerta Brasil (10), que têm atingido bons resultados na recuperação de veículos. Contudo, muitas vezes, os veículos só são recuperados após o contraventor cometer outros crimes utilizando estes veículos como meio de transporte e fuga (11).

Atualmente, ao melhor de nosso conhecimento, apesar de haver algumas técnicas e pesquisas, não existe uma ferramenta na mão das Polícias para auxiliar na predição de rota veicular de veículos roubado, deixando os Batalhões Policiais com uma vastidão territorial de possibilidades. Uma ferramenta deste tipo auxiliará o combate a crimes, evitando migrações de criminalidade indesejadas como ao utilizar policiamento ostensivo, aquele visível e explícito para reprimir possíveis crimes locais.

Assim, utilizando informações existentes no âmbito do Ministério da Justiça e Segurança Pública, como as presentes na Diretoria de Inteligência – DINT e no Centro Integrado de Comando e Controle Nacional – CICCEN, e tecnologias como reconhecimento ótico de caracteres (Optical Character Recognition - OCR) que as câmeras de trânsito modernas e pedágios já possuem para leitura das placas de carro, associadas à geolocalização destas câmeras para localização momentânea dos veículos, estima-se a rota que os veículos roubados ou furtados irão traçar, auxiliando a polícia a interceptar estes veículos antes da ocorrência de outros crimes associados, como sequestros, roubos a banco, entre outros.

Nesse contexto, quando um veículo é roubado, o proprietário deve realizar um Boletim de Ocorrência (BO) em uma delegacia física ou virtual. Essa ação gera um alerta que, quando uma câmera de vigilância identifica por sua placa o veículo roubado, usando OCR - *Optical Character Recognition* - para traduzir a imagem da placa do veículo capturada pela câmera em texto, são registrados, também, o local e a hora

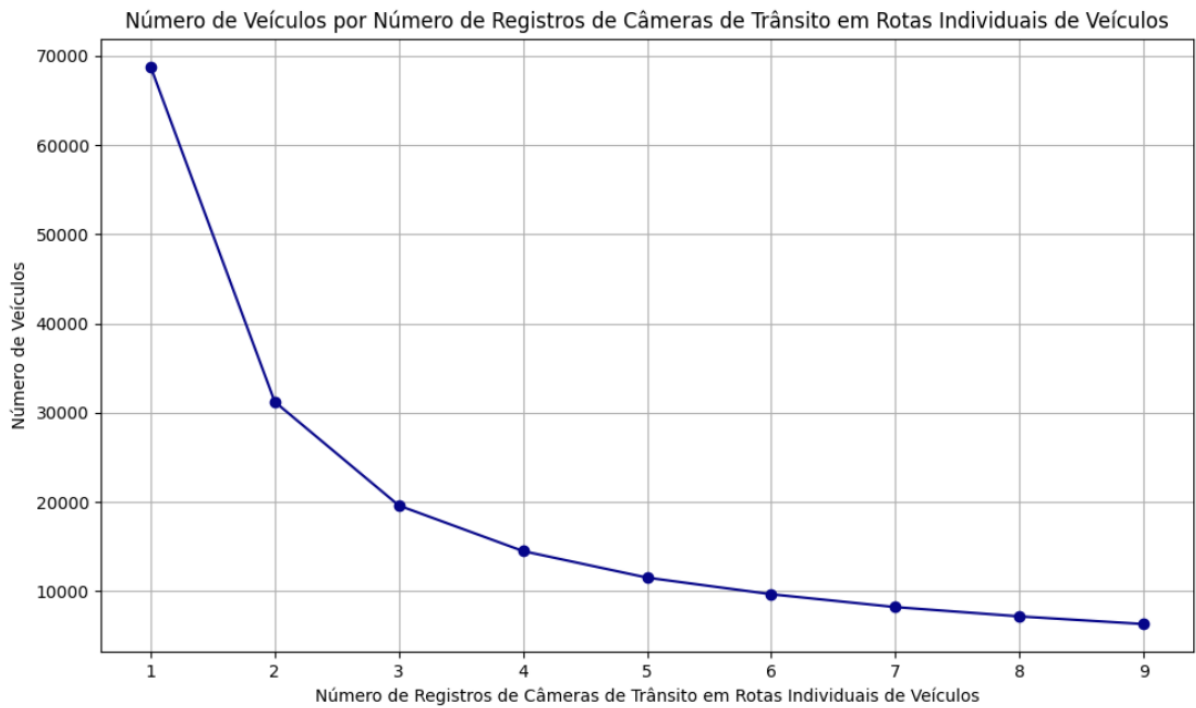


Figura 1.1: Veículos roubados em São Paulo, Brasil.

em que foi identificada a placa. Assim, usando radares de velocidade com Reconhecimento Óptico de Caracteres (OCR), conforme mencionado em (12), e conhecendo as geolocalizações desses equipamentos, é possível identificar a localização geográfica momentânea de um veículo roubado, mesmo que este não possua nenhum equipamento de georreferenciamento anexado ou em seu interior. Isso permite a identificação de pontos discretos e esparsos ao longo da rota do veículo.

Nesse contexto, utilizar pontos históricos para prever o próximo sensor externo (radar de velocidade) onde um veículo roubado tem maior probabilidade de ser detectado pode auxiliar tanto na recuperação do veículo quanto na redução de crimes subsequentes que podem ocorrer após o roubo do veículo, conforme mencionado em (13).

No entanto, este cenário apresenta um dilema: utilizar o menor número possível de pontos de sequência histórica para prever e agir mais rapidamente na tentativa de recuperar o veículo roubado, ou aguardar por mais pontos de sequência para melhorar a precisão da próxima previsão de geolocalização, o que pode nunca ocorrer, já que veículos roubados tendem a possuir apenas alguns registros.

Para ilustrar essa questão, que é ainda mais complexa devido ao fato de que um condutor agindo fora da lei naturalmente tende a apresentar um comportamento de direção anômalo em comparação com um cidadão típico em seu dia a dia (14), a Figura 1.1 mostra o número de veículos roubados em relação ao número de registros na base de dados anonimizados do estado de São Paulo (SP), Brasil. Por exemplo, um veículo com 3 registros também possui 2 ou 1 registro, mas não 4 registros. Assim, veículos com pelo menos 9 registros representam apenas 9,2% do total de veículos roubados com pelo menos 1 registro. Ou seja, quanto maior a necessidade de uma sequência histórica mais longa como entrada em um modelo de predição, menos veículos podem ser localizados.

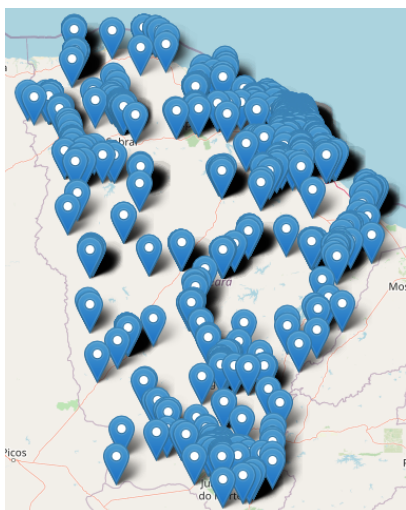


Figura 1.3: Sensores do CE.

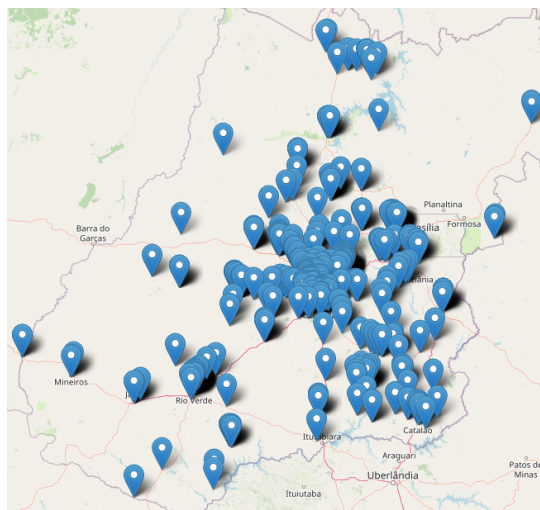


Figura 1.4: Sensores de GO.

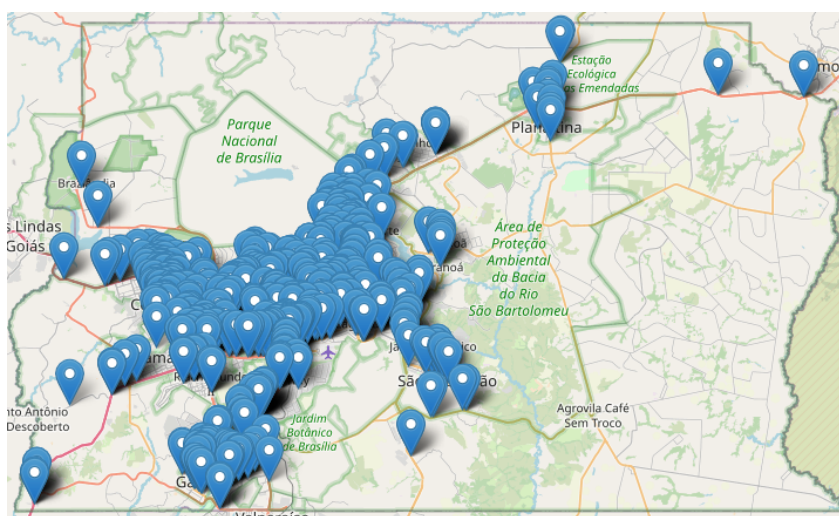


Figura 1.5: Sensores do DF.

(18), além de trabalhos como (19) e (15) voltados para a predição de trajetórias veiculares de veículos roubados e de táxis.

Uma questão que normalmente é colocada é que, apesar da arquitetura LSTM envolver memória de longo prazo (20), haveria uma desvantagem para sequências muito longas em relação ao Transformer (21). No entanto, para o contexto deste trabalho, isso não é tão relevante, pois as sequências para veículos roubados costumam ser curtas. Em grande parte, o veículo deixa de ser identificado porque foi abandonado, desmontado, placa adulterada ou, a melhor das opções, foi recuperado e parou a geração de alertas.

Assim, outros estudos têm abordado essa situação e a arquitetura de aprendizado de máquina Long-Short Term Memory (LSTM) tem sido uma das mais amplamente utilizadas em estudos de previsão de trajetória. Em alguns casos, mostrou-se ser a melhor, como em (22) e (23), enquanto em outros não, como em (24) e (25). Ainda assim, uma implementação específica da arquitetura Gradient Boosted Regression Trees (GBRT), também conhecida como Extreme Gradient Boosting (XGBoost), apresentou os melhores resultados em (13). Além disso, uma nova arquitetura chamada Transformer e LSTM Entrelaçados (TLE)

superou as demais ao usar apenas dois pontos de geolocalização para prever o terceiro ponto, conforme mostrado em (26). Como este estudo pretende analisar arquiteturas para previsão de sequência com poucos pontos históricos, ele se concentrará nessas quatro principais arquiteturas de ponta: LSTM, Transformer, TLE e XGBoost.

Dado, ainda, que o estudo (13) demonstrou desempenho superior para XGBoost em relação ao LSTM, e que o LSTM superou o TLE com sequências mais longas, mas apresentou desempenho inferior ao TLE com sequências mais curtas em (26), enquanto se considera também a influência das mudanças de representação de geolocalização destacadas em (13), este estudo busca explorar, adicionalmente, como essas descobertas se aplicam em condições diversas. De forma mais direta, nosso primeiro estudo objetivou atingir um ganho de predição, por meio das lacunas de pesquisa da inclusão de uma nova informação sobre a via da rua, ou seja, sobre qual trecho ou parte de toda a rua o veículo está transitando em determinado momento, até então não observada em outros estudos, no melhor do nosso conhecimento, e uma nova arquitetura entrelaçada com potenciais inovadores, enquanto nosso segundo estudo objetivou apoiar a definição de melhor arquitetura, por meio das lacunas de pesquisa de análises dos pontos de transição entre arquiteturas de melhor desempenho e do impacto da mudança no formato de representação geográfica, com potenciais de uma nova visão sobre como abordar essa situação. Especificamente, observando diferentes comprimentos de sequência e representações geográficas, buscamos responder às seguintes questões de pesquisa (QP) nos respectivos estudos:

1. Primeiro Estudo - Ganho de Predição:

- **QP1:** A informação sobre qual via da rua o veículo foi identificado auxilia na predição da próxima posição?
- **QP2:** A nova arquitetura Transformer e LSTM Entrelaçados (TLE) trouxe ganho em relação às arquiteturas individuais de LSTM e Transformer?

2. Segundo Estudo - Apoio à Definição de Melhor Arquitetura:

- **QP3:** Qual arquitetura apresenta melhor desempenho com comprimentos de sequência curtos para predição?
- **QP4:** Existe um ponto de transição onde se altera qual a arquitetura com melhor desempenho?
- **QP5:** O formato da representação geográfica impacta os resultados da predição?
- **QP6:** O formato da representação geográfica pode alterar o ponto de transição entre as arquiteturas com melhor desempenho, caso exista?
- **QP7:** A nova arquitetura Transformer e LSTM Entrelaçados (TLE) é vantajosa?

1.1 OBJETIVOS

Este trabalho tem como objetivo geral desenvolver uma arquitetura de predição de próximo sensor externo para geolocalização veicular ao se buscar a posição futura de um determinado veículo, identificado

por sua placa em câmeras com OCR (*Optical Character Recognition*), como em radares de velocidade, em contexto de veículo roubado ou furtado.

Baseado no objetivo geral, tem-se os seguintes objetivos específicos:

- Mapear os pontos geolocalizados onde é possível ter acesso à identificação dos veículos;
- Avaliar o estado-da-arte em arquiteturas de predição, em especial as aplicadas à predição de trajetórias e rotas veiculares;
- Aplicar as principais arquiteturas de predição ao conjunto de dados veiculares obtidos das câmeras;
- Avaliar os resultados das arquiteturas de predição aplicadas;
- Desenvolver uma nova arquitetura aplicada à solução de predição de próximo sensor externo tendo como base as outras arquiteturas que obtiverem melhores resultados;
- Refinar a arquitetura com base nos resultados aplicados em dados reais.

1.2 CONTRIBUIÇÕES DO TRABALHO

Ao observar demais trabalhos que envolvem Predição do Próximo Sensor Externo da Trajetória (*External Sensor Trajectory* (EST) Prediction, em inglês) que envolvem dados de veículos roubados, como em (13), (19), (27) e (28), este trabalho traz as seguintes contribuições específicas e seus diferenciais:

1. Análise do impacto na predição ao utilizar a informação sobre qual via da rua o veículo foi identificado. Ao se realizar predições, existe uma infinidade de variáveis que podem ser incluídas como entrada nos modelos de IA. Portanto, encontrar novas variáveis com dados disponíveis e obter resultados que impactam positivamente é sempre uma excelente contribuição. Melhor que isso é, como neste caso, em que a variável não exige grande aumento de processamento computacional para gerar o ganho de predição, obtendo uma notável relação de custo-benefício.
2. Análise em relação à transição de qual arquitetura obtém melhores resultados ao alterar o tamanho da sequência histórica para predição. Quando se realiza predições baseadas em uma sequência histórica como entrada no modelo de IA, definir o tamanho da sequência histórica é uma dúvida comum. Primeiro, porque quanto maior o tamanho da sequência, mais dados são necessários e, consequentemente, maior tempo de treinamento, assim como maior capacidade de armazenamento e de processamento. Segundo, pois nem sempre é uma relação linear entre o tamanho de sequência histórica e a melhoria na predição. Em outras palavras, nem sempre aumentar o tamanho da sequência de entrada do modelo de IA vai trazer benefícios. No caso específico de pontos de geolocalização de veículos roubados com sensores externos a situação ainda tem alguns agravantes, haja vista haver a propensão de empregarem o veículo em outros crimes subsequentes e a tendência de interrupção de novos registros de geolocalização dessa ocorrência, inviabilizando entradas e predições longas. Por conseguinte, uma análise de variação do tamanho de sequências históricas para valores curtos,

obtendo a melhor arquitetura de predição para cada um dos valores é de extrema relevância para otimização e apoio à decisão dos recursos envolvidos nessas predições.

3. Impacto do formato de representação geográfica no ponto de transição entre melhores arquiteturas. Ao contrário da crença de algumas pessoas, a Terra não é plana, o que dificulta e gera possíveis distorções na representação de uma localização geográfica. Para tratar essa complexidade, existem diferentes formas de representar uma geolocalização, cada uma com suas peculiaridades. Isso significa que as relações entre os pontos geográficos podem variar no contexto dos formatos em que são representados, posto isto, um modelo de IA para predição de geolocalização pode se beneficiar a depender do tipo de representação geográfica que é aplicada nas coordenadas dos veículos. Outrossim, esse ganho pode impactar de maneira diferente em cada tipo arquitetura de predição causando uma alteração no ponto de transição entre melhores arquiteturas em relação ao tamanho da sequência de entrada dos modelos.
4. Nova arquitetura utilizando o conceito de entrelaçamento entre Transformer e LSTM, criando a arquitetura Transformer e LSTM Entrelaçados (TLE). Observando os principais estudos relacionados à predição de geolocalização de veículos, há duas arquiteturas que se destacam e se alternam na posição dos resultados de melhor arquitetura, que são a LSTM e a Transformer. Esse panorama nos leva a crer que cada uma tem propriedades importantes no quesito de predição de trajetória veicular. Alguns estudos ainda tentaram algumas formas de junção das duas, como a arquitetura MHA-LSTM, porém sem mostrar vantagem conclusiva em relação a alguma das outras principais de forma individualizada. A forma inovadora de combinação das duas arquiteturas LSTM e Transformer utilizando um entrelaçamento e identificando a janela de aplicação à qual essa construção melhor se adapta torna-se uma contribuição de extrema relevância.

As nossas contribuições específicas corroboram para uma melhoria mais ampla da sociedade, logo, considerando ainda as contribuições em seu aspecto mais abrangente, podemos listar as seguintes:

1. A Sociedade Brasileira de Computação (SBC) trouxe em novembro de 2024 os Grandes Desafios da Computação no Brasil para os próximos dez anos (2025-2035) (29). Dentre esses desafios, a seguir delineamos como este trabalho contribui com três deles.
 - a) Inteligência Artificial na categoria dos Desafios Tecnológicos. Ao criar uma nova arquitetura de inteligência artificial de alta complexidade, usando *deep learning*, com uma capacidade inovadora, este trabalho contribui para uma aplicação real e ainda com potencial inovador que pode ser explorado em diversos outros cenários, como acontece com as arquiteturas primárias que embasaram a criação da Transformer e LSTM Entrelaçados (TLE).
 - b) Ubiquidade da Internet na categoria dos Desafios Tecnológicos. Este trabalho contribui com o avanço e a importância de múltiplas tecnologias conectadas via internet. Para a predição de próximo sensor externo, existe uma colaboração e conexão das câmeras de radares de velocidade que capturam imagens e identificam placas de veículos com base em OCR, enviando a informação a bases de dados remotas, que processam e realizam cruzamentos com outras bases de dados de alertas de veículos roubados para então, aplicando inteligência artificial, realizar predições de próximas localizações desses veículos e enviando às autoridades competentes.

- c) Outros Impactos Sociais e Econômicos da Computação na categoria dos Desafios Socioeconômicos. O avanço na recuperação de veículos roubados pode gerar impactos indiretos como a baixa no custo de seguro de veículos, maior eficiência da aplicação dos recursos públicos contra esses crimes, a redução de custos com perdas materiais dos cidadãos e a mitigação de impactos psicológicos derivados dessas transgressões.
2. Contribuição com o Objetivo de Desenvolvimento Sustentável (ODS) 16 da Organização das Nações Unidas (ONU), ODS-16. O Brasil é um dos 51 países fundadores da ONU em 1945. A ONU tem atualmente 17 Objetivos de Desenvolvimento Sustentável vinculados à Agenda 2030, sendo o ODS-16 a “Paz, Justiça e Instituições Eficazes”, que traz em seu item 16.4 o texto *"Até 2030, reduzir significativamente os fluxos financeiros e de armas ilegais, **reforçar a recuperação e devolução de recursos roubados** e combater todas as formas de crime organizado"* (ONU, Agenda 2030, ODS-16, item 16.4, grifo nosso) (30). Este trabalho contribui justamente com o aspecto de recuperação e devolução de veículos roubados e evitar crimes subsecutivos ao roubo desses veículos, bem como melhorando o bem-estar social e o sentimento de segurança para a sociedade.

1.3 PUBLICAÇÕES

Os resultados obtidos reforçam a relevância das análises realizadas e indicam potenciais avanços no estado da arte da área estudada. As contribuições acadêmicas deste trabalho estão refletidas a seguir.

O artigo *"Predição de Geolocalização de Veículo com Alerta de Roubo Usando LSTM, Transformer e TLE"* (26), por meio de sensores externos - em geral câmeras de radares de velocidade -, trouxe a análise da inclusão da informação sobre qual via da rua um veículo roubado foi identificado como dado adicional, junto com as respectivas geolocalizações, nas sequências de entrada dos modelos de Inteligência Artificial (IA), demonstrando um impacto positivo nas predições. Ainda em destaque, trouxe a primeira aplicação da nova arquitetura de aprendizado profundo (*deep learning*) Transformer e LSTM Entrelaçados (TLE), mostrando potencial inovador para sequências históricas mais curtas como entrada em modelos de predição de próximo sensor externo:

- MACEDO, GUSTAVO V. I. DE ; ROCHA FILHO, GERALDO P. ; SANTOS, JOÃO K. M. DOS ; NEVES, ARTHUR R. ; ALMEIDA, MURILO G. ; FALQUEIRO, MARIANA C. ; MENEGUETTE, RODOLFO I. ; SERRANO, ANDRÉ L. M. ; MENDONÇA, FÁBIO L. L. DE ; GONÇALVES, VINÍCIUS P. . Predição de Geolocalização de Veículo com Alerta de Roubo Usando LSTM, Transformer e TLE. In: Simpósio Brasileiro de Computação Ubíqua e Pervasiva, 2024, Brasil. Anais do XVI Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP 2024). Porto Alegre: Sociedade Brasileira de Computação - SBC, 2024. v. 16. p. 61-70.
- Segundo Melhor Artigo do CSBC/SBCUP 2024 - XLIV Congresso da Sociedade Brasileira de Computação - XVI Simpósio Brasileiro de Computação Ubíqua e Pervasiva, Sociedade Brasileira de Computação - SBC.

Com a premiação do nosso primeiro artigo, fomos convidados a enviar uma versão estendida desse

artigo para a revista *Journal of Internet Services and Applications* (JISA). Assim, o artigo "*Influence of Sequence Length and Geographic Representation on Optimal Prediction Architectures for Stolen Vehicle Geolocation*" traz uma visão mais abrangente em relação a aspectos relevantes ao definir a melhor arquitetura a ser empregada na predição de próximo sensor externo para rotas veiculares de veículos roubados. Nesse sentido, observa-se que a depender do tamanho da sequência histórica, ou seja, quantos pontos serão utilizados para prever o próximo, pode-se obter uma predição melhor com uma ou com outra arquitetura, que nesse estudo a arquitetura TLE foi melhor para sequências menores e LSTM melhor para sequências maiores, corroborando com a tendência apresentada no estudo anterior. Outro ponto analisado foi o impacto da alteração do formato da representação geográfica, que além de ser capaz de gerar impacto positivo ou negativo aos resultados de predição, também desempenhou mudança no ponto de transição entre as melhores arquiteturas. Isso significa que alterar o formato de representação geográfica pode alterar qual a melhor arquitetura a ser empregada a depender do tamanho da sequência histórica definida:

- MACEDO, GUSTAVO V. I. DE ; ROCHA FILHO, GERALDO P. ; SANTOS, JOÃO K. M. DOS ; NEVES, ARTHUR R. ; ALMEIDA, MURILO G. ; FALQUEIRO, MARIANA C. ; MENEGUETTE, RODOLFO I. ; SERRANO, ANDRÉ L. M. ; MENDONÇA, FÁBIO L. L. DE ; GONÇALVES, VINÍCIUS P. *Influence of Sequence Length and Geographic Representation on Optimal Prediction Architectures for Stolen Vehicle Geolocation*. Revista *Journal of Internet Services and Applications* (JISA), 2025. (No prelo).

1.4 ESTRUTURA DA DISSERTAÇÃO

Este trabalho apresenta os demais capítulos conforme a seguinte organização: o Capítulo 2 apresenta os conceitos; o Capítulo 3 traz os trabalhos relacionados; o Capítulo 4 descreve a metodologia; o Capítulo 5 detalha a arquitetura TLE; o Capítulo 6 a aplicação da arquitetura TLE para os estados do DF, GO e CE; o Capítulo 7 a aplicação da arquitetura TLE para o estado de SP; o Capítulo 8 discute as conclusões, limitações e trabalhos futuros, seguidos pelas referências bibliográficas.

2 CONCEITOS

Ao contribuir com o mundo acadêmico é natural se aprofundar em ramos específicos do conhecimento de tal forma que, para uma melhor e mais eficiente compreensão do que será apresentado neste trabalho, é importante apresentarmos alguns conceitos, definições e detalhes relacionados ao tema deste trabalho e que auxiliarão no adequado entendimento daquilo que será abordado mais a frente. Desse modo, neste Capítulo trouxemos uma visão aprimorada da definição do problema, o modelo de funcionamento das arquiteturas de IA implementadas e alguns aspectos sobre três diferentes formatos geográficos utilizados neste trabalho para se referenciar à geolocalização de um ponto na Terra.

2.1 DEFINIÇÃO DO PROBLEMA

Para a predição da geolocalização de um veículo no contexto do próximo sensor externo, ou problema de *External Sensor Trajectory* (EST), existe uma margem inerente de imprecisão. Em ruas com múltiplas faixas, são registradas as mesmas coordenadas de latitude e longitude, independentemente da faixa em que o veículo foi identificado. Adicionalmente, para fins de predição e considerando como um problema de regressão para o modelo de Inteligência Artificial (IA), a geolocalização prevista pode não corresponder à posição de um sensor externo existente, sendo necessário um ajuste para alinhá-la às coordenadas de latitude e longitude de um sensor externo conhecido. Além disso, há a imprecisão na trajetória real completa, pois apenas posições pontuais e esparsas são registradas.

A Figura 2.1 apresenta uma ilustração de predição do próximo sensor externo. Nesta imagem, o veículo azul poderia virar à esquerda, onde seria identificado novamente pelo sensor externo na parte superior da figura. Alternativamente, ele poderia seguir em frente e ser identificado pelo sensor à direita da figura. Por fim, ele poderia virar à direita e evitar ser identificado por qualquer outro sensor externo próximo ao ponto inicial. No caso ilustrado na Figura 2.1, a predição conclui que o veículo provavelmente seguirá em frente. Esse cenário demonstra que o uso da informação da faixa de rua onde o veículo foi identificado pode auxiliar na predição do próximo sensor externo, mesmo sem o conhecimento exato da rota intermediária entre os pontos (26).

Nesse sentido, prever a trajetória do próximo sensor externo, devido aos dados esparsos e discretos, apresenta desafios maiores do que a predição de trajetória comum - quando se tem informações contínuas e completas sobre o trajeto prévio -, especialmente porque o caminho entre os sensores é desconhecido, mas, ainda assim, importante de ser representado. A **Figura 2.2** mostra um exemplo real de uma trajetória mapeada usando registros de sensores externos. Observa-se que o caminho real entre os pontos de geolocalização não pode ser definido, razão pela qual a trajetória do veículo é representada como uma linha reta conectando os pontos dos radares de velocidade em sequência, sem ilustrar a rota intermediária real.

Após compreender visualmente a situação do problema, podemos definir o problema de maneira mais precisa. Quando um veículo passa por um sensor externo capaz de identificá-lo, é gerado um registro $r = (p, s, t)$, onde p representa os dados de identificação única do veículo, s representa os dados de identi-

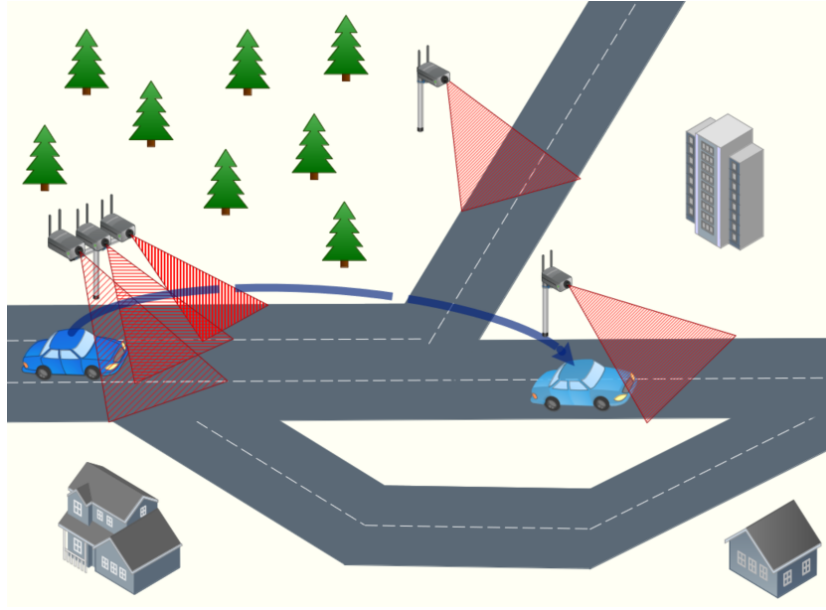


Figura 2.1: Ilustração da predição do próximo sensor externo.

ficação única do sensor e t representa o carimbo de tempo (*timestamp*) do registro. Cada sensor s registra passagens apenas em uma única faixa de rua. No entanto, cada localização geográfica, $geo = (lat, long)$, pode corresponder a um único sensor s ou a um conjunto de sensores $(s_1, s_2, \dots, s_n)$ localizados em diferentes faixas de rua na mesma área. Em outras palavras, cada sensor s tem uma localização geográfica correspondente, $geo = (lat, long)$, que pode ou não ser compartilhada com outros sensores s . Portanto, s pode ser substituído por geo , significando que $s = (lat, long)$. Os *timestamps* t são usados para garantir a ordem cronológica correta dos dados dos registros. Como mostrado por (26), a melhor abordagem é utilizar ambas as informações de localização como entrada: $r = (p, s, lat, long)$. Este trabalho adotará essa abordagem.

Para a predição, uma sequência de registros $seq = (r_1, r_2, \dots, r_n)$ é usada como entrada, com o objetivo de prever a localização real do próximo sensor externo r_{n+1} como saída. Portanto, cada modelo de arquitetura é treinado para encontrar uma função de regressão $f(x)$ tal que $f(seq) = r_{n+1}$.

Além disso, além da informação sobre a faixa de rua, outro fator que pode impactar os resultados da predição é como a geolocalização do sensor externo é representada. Por exemplo, para o mesmo ponto geográfico, em vez de usar latitude e longitude, é possível empregar *Northing* e *Easting*, potencialmente utilizando diferentes sistemas de referência. Isso ocorre porque os modelos de IA, especialmente em problemas de regressão como este, processam dados numéricos e, dependendo das relações inerentes a cada representação geográfica, bem como da distribuição e distâncias entre os registros de geolocalização, o modelo pode capturar padrões nas sequências de dados de entrada de forma melhor, igual ou pior.

Entendidas as peculiaridades dessa situação problema, na próxima seção apresentaremos o referencial teórico base para uma melhor compreensão deste trabalho.

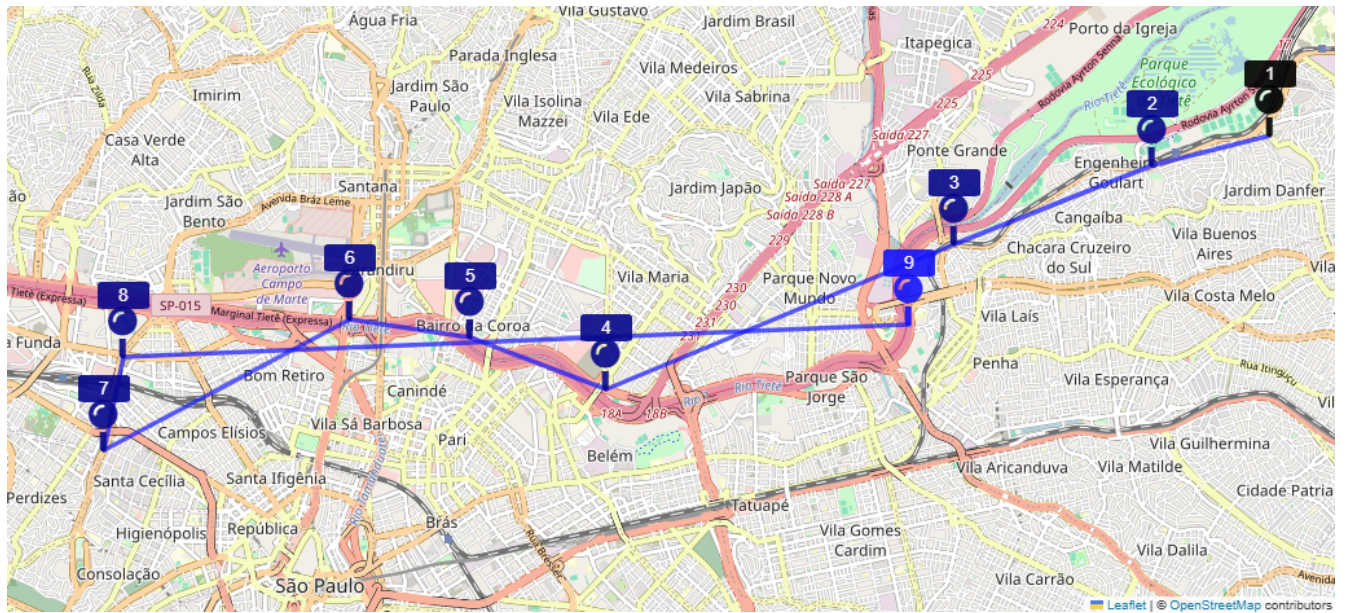


Figura 2.2: Exemplo de uma trajetória com 9 registros de sensores externos em SP.

2.2 REFERENCIAL TEÓRICO

2.2.1 Arquiteturas de Aprendizado de Máquina

Ao longo deste trabalho, quatro arquiteturas foram empregadas para realização dos treinamentos e das predições: XGBoost, LSTM, Transformer e Transformer e LSTM Entrelaçados (TLE). Neste capítulo serão apresentados os fundamentos necessários para se entender as arquiteturas XGBoost, LSTM e Transformer utilizadas no decorrer deste trabalho. A arquitetura Transformer e LSTM Entrelaçados (TLE), que se utiliza dos embasamentos da LSTM e da Transformer, será explicada e detalhada no Capítulo 5.

A XGBoost (*Extreme Gradient Boosting*) é uma implementação otimizada usando conjuntos de árvores de decisão, sendo categorizada como aprendizado de máquina (*machine learning*). Em outro ponto de vista, a LSTM, a Transformer e a TLE entram na categoria de aprendizado profundo (*deep learning*), que é uma subcategoria de aprendizado de máquina, pois baseiam-se em redes neurais profundas, ou seja, com múltiplas camadas, incluindo camadas ocultas (*hidden layers*) - camadas que não se conectam diretamente nem com os dados de entrada nem com os dados de saída do modelo de IA -, além de trabalhar com operações matriciais, funções de ativação diferenciáveis e retropropagação, que são operações mais complexas que as árvores de decisão (31) (32). A Figura 2.3 ilustra a relação dessas arquiteturas e suas categorias.

Neste trabalho, é aplicado o aprendizado supervisionado, que é quando se realiza o treinamento do modelo fornecendo respostas corretas para a predição requerida. Nesse contexto, é aplicada a função de perda (*Loss Function*) que, em termos resumidos, verifica o quanto o modelo errou em sua predição e, com base nisso, o modelo utiliza a retropropagação (*backpropagation*) para ajustar seus pesos e otimizar seus parâmetros para reduzir o erro em sua próxima predição. Esse é um processo iterativo que ocorre durante o treinamento do modelo, que visa, além de acertar a próxima predição, melhorar a generalização evitando a simplificação excessiva (*underfitting*) e o ajuste excessivo (*overfitting*) (33) (34).

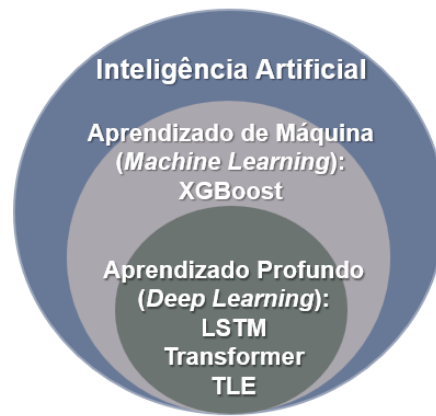


Figura 2.3: Relação entre subcategorias de inteligência artificial e arquiteturas deste trabalho.

Na sequência, serão apresentados mais detalhes sobre como funcionam cada uma das três arquiteturas que foram aplicadas em estudos ao longo deste trabalho - XGBoost, LSTM e Transformer -, deixando os detalhamentos da arquitetura TLE para o Capítulo 5.

2.2.1.1 Arquitetura XGBoost

A arquitetura XGBoost (Extreme Gradient Boosting) é uma implementação altamente eficiente e escalável de Árvores de Regressão com Gradiente Reforçado (*Gradient Boosted Regression Trees* - GBRT), projetada para dados estruturados, incluindo dados sequenciais. O modelo constrói um conjunto de árvores de decisão de forma sequencial, onde cada árvore visa corrigir os erros cometidos pelas anteriores. O XGBoost é eficaz para tarefas como classificação e regressão, sendo capaz de lidar com grandes conjuntos de dados e interações complexas de características com ótimo desempenho. Essa arquitetura se popularizou com o artigo (1) e a Figura 2.4 exemplifica seu funcionamento.

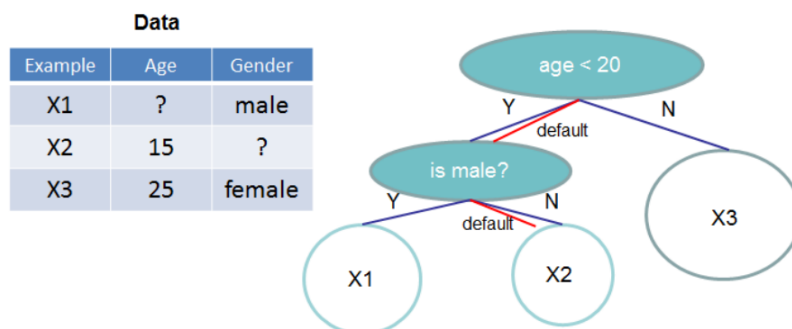


Figura 2.4: Funcionamento da arquitetura XGBOOST. Fonte: CHEN et al. (1).

A arquitetura XGBoost utilizada neste trabalho foi baseada na arquitetura GBRT descrita em (13), estudo no qual teve grande relevância. No entanto, por ser uma arquitetura mais simples e que não atinge papel de destaque nos resultados deste trabalho, dispensa maiores aprofundamentos teóricos.

2.2.1.2 Arquitetura LSTM

A arquitetura *Long Short-Term Memory* (LSTM) é um tipo de rede neural recorrente (*Recurrent Neural Network* - RNN) e foi especificamente projetada para enfrentar os desafios do aprendizado sequencial. A principal característica da arquitetura LSTM é sua capacidade de “lembrar” informações ao longo de grandes sequências, permitindo capturar dependências temporais e padrões no decorrer do tempo. A forma como foi projetada também evita o Problema da Dissipação do Gradiente (*Vanishing Gradient Problem*) (35) (36). Outra vantagem é que ao processar os dados passo a passo, ela garante que a ordem dos eventos seja intrinsecamente considerada durante o aprendizado, isso a torna altamente eficaz para tarefas que envolvem dados temporais, como séries temporais e previsões de trajetórias, como é o caso deste trabalho.

A Figura 2.5 apresenta o funcionamento interno da arquitetura LSTM, onde X_t são os dados de entrada e h_t os dados de saída de cada célula LSTM. As células LSTM intermediárias são conectadas de forma recorrente, onde suas entradas são, além dos dados X_t , as saídas da célula anterior h_{t-1} trazendo informações de curto prazo e o contexto C_{t-1} trazendo as informações de longo prazo. Tanto as informações de curto prazo h_{t-1} quanto as de longo prazo C_{t-1} passam por cálculos internos dentro da célula antes de se tornarem as saídas C_t e h_t , as quais servirão de dados de entrada, juntamente com X_{t+1} , na célula seguinte, e assim sucessivamente. Nessa imagem, tanto a função sigmoide (σ) quanto função tangente hiperbólica ($\tanh$) são funções de ativação usadas dentro da célula LSTM, tendo sigmoide com valores entre 0 e 1 atuando como *gates* (portas) para regular a passagem de informação, enquanto a tangente hiperbólica com valores entre -1 e 1 regulando os valores do estado da célula e da saída. Essas informações e o processo de recorrência também é representado na Figura 2.5 e detalhado em (2) e (37),

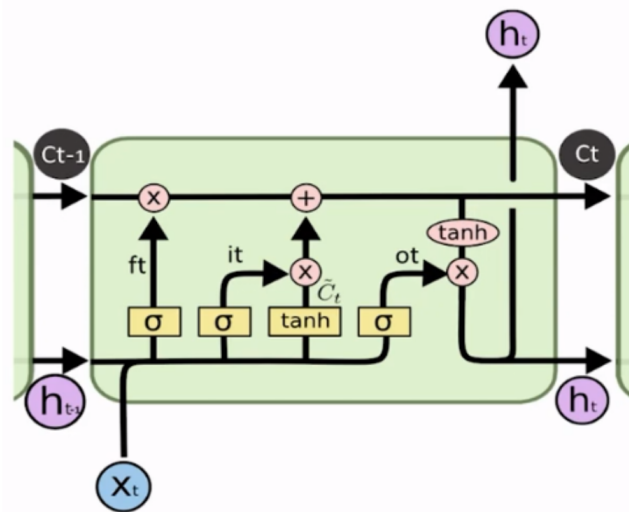


Figura 2.5: Funcionamento da arquitetura LSTM. Fonte: TECH (2).

Essa característica de considerar informações de curto e longo prazo, sendo capaz de “lembrar” ou “esquecer” aprendizados históricos, torna a LSTM uma das arquiteturas mais utilizadas em aprendizado sequencial e com resultados de destaque em diversos artigos. Neste trabalho não foi diferente, ela mostra ter um papel importante em diversos resultados, inclusive na criação da nova arquitetura Transformer e LSTM Entrelaçados (TLE).

2.2.1.3 Arquitetura Transformer

Em uma visão diferenciada, a arquitetura Transformer revolucionou a área de Processamento de Linguagem Natural (PLN) ao trazer a abordagem de “atenção” às palavras mais importantes em uma sequência de palavras e conseguir fazer isso em um processamento paralelo, em vez de um processamento sequencial como acontece na arquitetura *Long Short-Term Memory* (LSTM) e outras arquiteturas de *Recurrent Neural Network* (RNN) como a *Gated Recurrent Units* (GRU). O conceito de “atenção” já havia sido proposto anteriormente (38), entretanto, o artigo “Attention Is All You Need” (3) tornou-o o mecanismo principal da arquitetura, possibilitando que cada elemento da sequência “preste atenção” em todos os demais elementos simultaneamente, dispensando o processamento etapa a etapa típico das RNNs, como era utilizado na técnica de (38). O artigo (3) trouxe essa abordagem de forma paralelizada e apresentou a Figura 2.6 para ilustrar o funcionamento da arquitetura Transformer, onde o grupo de funcionalidades à esquerda atua como *encoder* (codificador) processando a sequência de entrada e gerando representações contextuais. E o grupo à direita atua como *decoder* (decodificador) utilizando as representações para gerar as saídas, tendo cada bloco desses repetido por 6 vezes no artigo original, porém podendo ser ajustado para outros valores (Nx). Por fim, a camada Linear projeta os vetores de saída para a dimensão do vocabulário e a Softmax converte os valores da última camada em probabilidades de predição representando a probabilidade de cada palavra do vocabulário ser a próxima palavra gerada. Essa arquitetura atua bem paralelizada, substituindo recorrência por mecanismos de atenção multi-cabeças, conexão residual e normalização.

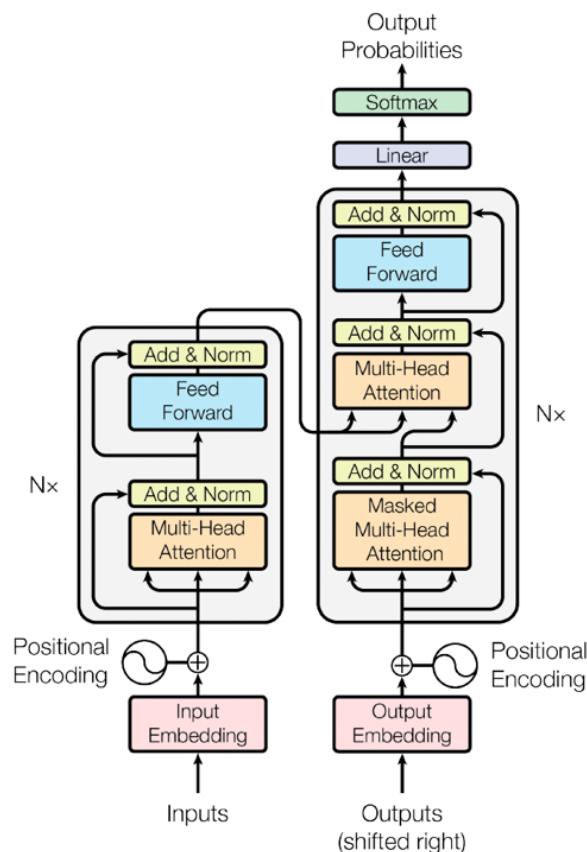


Figura 2.6: Funcionamento da arquitetura Transformer. Fonte: VASWANI et al. (3).

Nesse ponto, precisamos lembrar que computadores não entendem palavras como os seres humanos.

Máquinas “entendem” números, e, portanto, a arquitetura Transformer, que foi projetada para lidar com sequências de palavras, possui mecanismos para traduzir as palavras em números. Neste trabalho, considerando que estamos utilizando números diretamente e abordando a situação como um problema de regressão, alguns desses mecanismos não foram necessários e, por conseguinte, foram retirados, durante sua reprodução base para a utilização dessa arquitetura, para esse contexto, sendo eles: *Positional Encoding*, *Output Embedding* e *Softmax*.

Os componentes *Positional Encoding*, *Output Embedding* e *Softmax* foram concebidos originalmente para lidar com *tokens* textuais e gerar distribuições de probabilidade na saída, aspectos específicos para o processamento de linguagem natural. Neste cenário de problema de regressão com dados numéricos de entrada e saída, esses componentes se tornam desnecessários e, por essa razão, foram removidos. Assim, por se destacar em predição de sequências, incluindo de trajetórias, como nos estudos (25), (24) e (15), essa arquitetura Transformer adaptada ao contexto deste trabalho foi utilizada como um dos embasamentos para a criação da nova arquitetura TLE.

2.2.2 Representação Geográfica

Diferentemente de alterar a unidade de medida de metros para pés, por exemplo, onde continua se medindo da mesma forma, na mesma direção, modificando apenas o tamanho da unidade de medida, a forma como se realiza a representação geográfica altera não apenas o tamanho da unidade de medida, mas também a forma como se medem as distâncias. Isso pode ocasionar impacto na predição de geolocalizações.

Ao avaliar uma arquitetura para predição de geolocalização, naturalmente a forma como se mede e se representa essa medida geográfica, ao passar por um modelo de Inteligência Artificial - IA de predição da próxima geolocalização, pode gerar impactos positivos ou negativos importantes de serem analisados. No entanto, essas alterações precisam seguir algum padrão para manter a relação entre a posição real e sua representação numérica ou no mapa e que seja compreendida por outros indivíduos.

O European Petroleum Survey Group (EPSG) era o nome original da organização responsável por estabelecer um conjunto padronizado de sistemas de referência de coordenadas. Cada código EPSG serve como um identificador único usado para definir um sistema de referência de coordenadas específico. Esses códigos especificam como os dados geográficos são projetados em uma superfície plana, como um mapa, garantindo que diferentes conjuntos de dados se alinhem corretamente quando combinados (39). As três representações geográficas utilizadas neste trabalho são as seguintes:

- **EPSG:4326**, também conhecida como Latitude e Longitude na versão mais recente do Sistema Geodésico Mundial de 1984 (World Geodetic System 1984 - WGS84), que utiliza coordenadas geográficas de latitude e longitude em graus. A latitude geodésica é o ângulo contado sobre o meridiano que passa por P, compreendido entre a normal passante por P e o plano equatorial. Já a longitude geodésica é o ângulo contado sobre o plano equatorial, compreendido entre o meridiano de Greenwich e o ponto P. A representação de um ponto P é ilustrada na Figura 2.7. As bases de dados acessadas para os experimentos deste trabalho contêm por padrão as geolocalizações representadas neste formato EPSG:4326 (WGS84), que é amplamente difundido e conhecido por representar as coordenadas como latitude e longitude. Em consequência, a maior parte dos experimentos deste

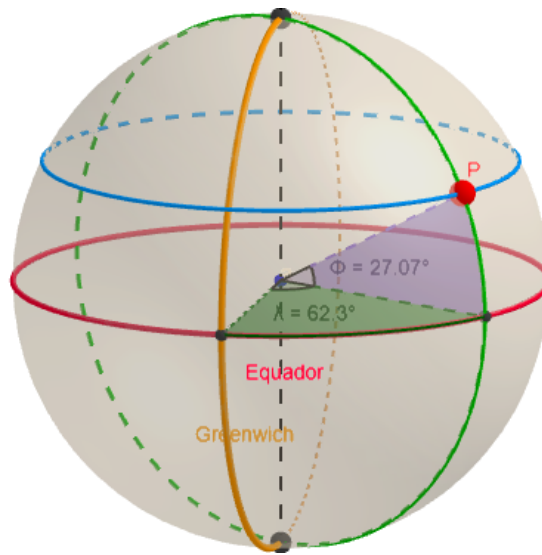


Figura 2.7: Exemplo da representação geográfica de um ponto P utilizando EPSG:4326 (WGS84). Fonte: MATHIAS et al. (4).

trabalho utilizaram esse formato.

- **EPSG:3857**, também conhecida como Web Mercator, que utiliza *Easting* e *Northing* medidos em metros, em vez de graus como em latitude e longitude. Trata-se de uma projeção cilíndrica tangenciando a linha Equatorial, considerando a Terra como uma esfera e projetando até 85,05° de latitude tanto para o norte quanto para o sul em um plano. Esse formato de representação geográfica, apesar de ser muito utilizado, não apresentou relevância significativa neste trabalho e, portanto, não se faz necessário maior aprofundamento técnico. A Figura 2.8 ilustra essa projeção cilíndrica em um plano. Apesar de não ter atingido destaque neste trabalho, esse formato EPSG:3857 (Web Mercator)

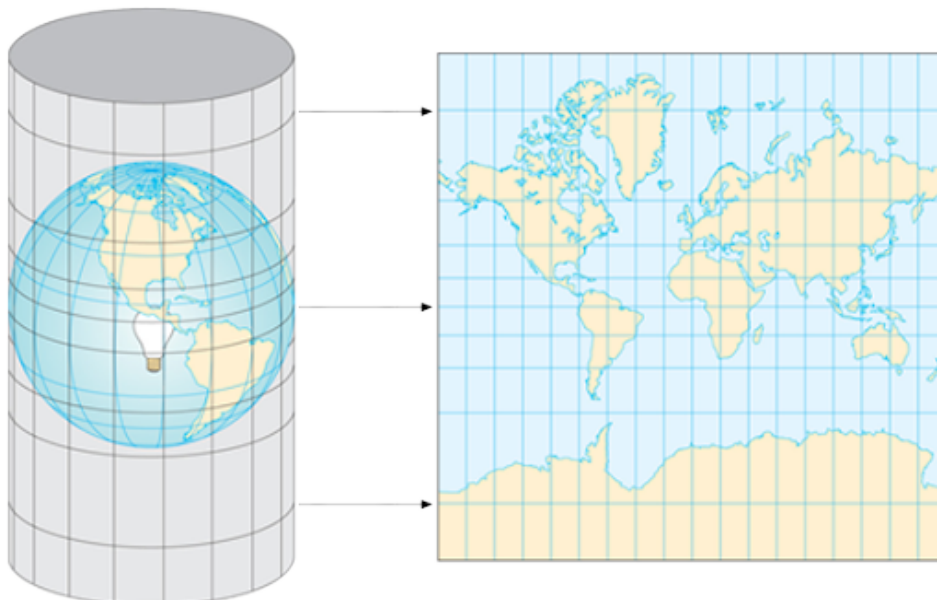


Figura 2.8: Exemplo de projeção cilíndrica em um mapa plano utilizando EPSG:3857 (Web Mercator). Fonte: GURGEL (5).

foi considerado para análise em vista de ser amplamente utilizado em soluções como Google Maps, Bing Maps, entre outros, pois permite projetar a superfície da Terra em uma tela 2D de maneira que seja fácil de visualizar em navegadores web e com melhor performance e compatibilidade ao aplicar zoom pelo usuário. Essas vantagens seriam úteis para o desenvolvimento de uma aplicação web baseada nos resultados deste trabalho, caso os resultados de predição fossem iguais ou melhores que em outros formatos.

- **EPSG:32723**, também conhecida como Universal Transverse Mercator (UTM), que também utiliza *Easting* e *Northing* medidos em metros, mas com diferentes sistemas de referência. É um sistema de projeção cilíndrica transversa, e considera que o cilindro gira em volta da Terra, realizando cortes secantes para cada zona de projeção, dividindo a Terra em 60 zonas de 6° cada, com a contagem iniciando em 180° de longitude (anti-meridiano de Greenwich). Também é uma projeção conforme, pois os ângulos e direções são preservados na projeção. Considera como referência horizontal um Meridiano Central em 500.000 metros, que diminui para Oeste e aumenta para Leste. Como referência vertical a linha do Equador tem dois valores: zero metros para o hemisfério Norte, aumentando para Norte até 10 milhões de metros (na latitude de 80°N); e como 10 milhões de metros para o hemisférios sul, reduzindo para o Sul até atingir zero (na latitude de 84°S). A Figura 2.9 ilustra essa divisão de zonas e hemisférios. Esse formato é amplamente utilizado em projetos de engenharia e

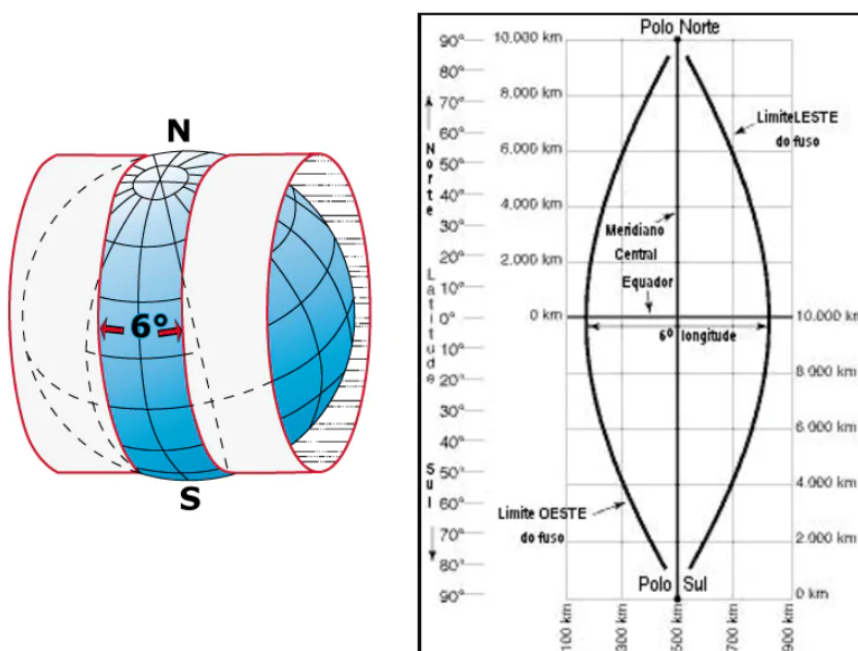


Figura 2.9: Exemplo de recorte de zonas e hemisférios utilizando EPSG:32723 (UTM). Fonte: GIOVANINI (6).

cartografia, pois oferece maior precisão em áreas locais, como estados e municípios, além de ser compatível com o WGS84. Por ser baseado em projeções por zonas, o EPSG:32723 (UTM), reduz distorções dentro de cada zona, proporcionando medições precisas de distâncias e áreas. Devido a essas características, foi selecionado para compor os experimentos deste trabalho, que considera trajetórias veiculares em áreas limitadas do Brasil.

Os três sistemas apresentados de representação geográfica EPSG:4326 (WGS84), EPSG:3857 (Web Mercator) e EPSG:32723 (UTM) diferem em como projetam a superfície da Terra, o que pode impactar as predições de geolocalização de veículos. O EPSG:4326 (WGS84) representa as localizações usando latitude e longitude em graus, assim como os dados originais deste trabalho, sendo ideal para cobertura global, porém menos preciso para predições locais devido à curvatura da Terra. O EPSG:3857 (Web Mercator) projeta as coordenadas geográficas em uma representação plana usando metros, o que distorce as distâncias, especialmente próximo aos polos, porém é comumente utilizado para mapeamento na web, exibindo dados geográficos em mapas digitais interativos bidimensionais para visualizar informações geográficas, características que seriam úteis na implementação de soluções web baseadas neste trabalho. Já o EPSG:32723 (UTM) também utiliza metros, contudo divide a Terra em zonas, proporcionando medições locais mais precisas, tornando-o mais adequado para predições em escala regional, como nos experimentos deste trabalho. Essas diferenças em precisão e projeção podem influenciar a precisão das predições de geolocalização de veículos (40)(41)(42).

Embora este estudo utilize três diferentes representações geográficas, elas compartilham uma estrutura comum: dois componentes de informação são utilizados, um para a distância horizontal, seja Longitude ou *Easting*, e outro para a distância vertical, seja Latitude ou *Northing*. Notavelmente, *Easting* é tipicamente mencionado primeiro, pois, em sistemas de coordenadas projetados, segue-se a convenção de tratar *Easting* como a coordenada x e *Northing* como a coordenada y, seguindo o formato (x, y). Em contraste, para Latitude e Longitude, as coordenadas geográficas são geralmente listadas como (Latitude, Longitude), com latitude aparecendo primeiro, em conformidade com a convenção geográfica de (Latitude, Longitude). Essa característica de estrutura comum desses formatos proporcionou menor necessidade de adaptação dos fluxos de dados ao converter os tipos de representação geográfica dos pontos de geolocalizações.

2.3 CONSIDERAÇÕES FINAIS

Ao definir o problema de forma visual e textual, podemos ter maior clareza e entendimento do contexto, das análises e dos resultados apresentados ao longo deste trabalho. Entendido o problema, importante entendermos as opções de como abordá-lo, nesse sentido, foram apresentados os funcionamentos internos de três arquiteturas (XGBoost, LSTM e Transformer) comuns para predição de trajetória veicular, que obtiveram relevância em outros estudos e foram utilizadas também neste trabalho. Destacam-se, ainda, duas delas, LSTM e Transformer, que foram utilizadas como base para a nova arquitetura TLE que será apresentada no Capítulo 5. Por fim, as descrições e imagens auxiliaram o entendimento de algumas características e peculiaridades de três diferentes formatos de representação geográfica, que se tornam expressivos no contexto de predição de geolocalizações de trajetória veicular, deixando mais claro o porquê da mudança no formato de uma representação geográfica poder impactar no desempenho das predições dos modelos de IA. No Capítulo 3 serão apresentados os trabalhos relacionados e como este trabalho se diferencia deles.

3 TRABALHOS RELACIONADOS

Predizer a trajetória do próximo sensor externo, sobretudo quando não há informações sobre o trajeto percorrido entre os sensores, se torna mais desafiador do que em cenários de dados contínuos. Apesar das diferenças nos tipos de dados utilizados nos experimentos, algumas abordagens gerais podem ser observadas e aplicadas de maneira similar em diferentes contextos. Nesse sentido, os estudos foram agrupados em duas categorias: inclusão de novas informações para predição; e nova arquitetura de predição.

Na categoria de inclusão de novas informações para predição são agrupados os artigos que incluíram novos dados, além das coordenadas básicas, juntamente com a sequência de entrada nos modelos de IA para predição de coordenadas futuras. Já na categoria de nova arquitetura de predição são agrupados os artigos em que naquele momento foram utilizadas arquiteturas inovadoras para a abordagem do problema de predição de coordenada futura considerando sequências de entrada.

3.1 INCLUSÃO DE NOVAS INFORMAÇÕES PARA A PREDIÇÃO

Em (23), o autor considera importante avaliar modelos no estado-da-arte no contexto de predição de trajetória de veículo balístico não manobrável. O estudo compara predições de trajetórias de voo de veículos balísticos, que, por ser uma trajetória de voo, é incluído um terceiro valor, a coordenada z em adição às coordenadas x e y . Essa dinâmica adiciona uma nova camada de complexidade à predição. Nesse ponto é importante destacar que para um modelo de IA, por não compreender o mundo real como um ser humano, trata-se apenas de mais uma variável a ser treinada e prevista. O estudo compara arquiteturas de *Long Short-Term Memory* (LSTM), *Gated Recurrent Unit* (GRU), *Transformer* e *Multi-Layer Perceptron* (MLP), com a LSTM obtendo os melhores resultados. Interessante analisar que o estudo utiliza 30 posições como entrada para prever 45 posições de saída. Naturalmente, quanto maior a distância da predição, maior será o erro, independentemente da arquitetura utilizada. Contudo, o fato de utilizar uma grande quantidade de pontos de entrada (trinta) preconiza o resultado da LSTM atingir melhor desempenho, o que coaduna com os resultados deste trabalho, como será observado adiante. No entanto, o que este trabalho traz é uma arquitetura *Transformer* e *LSTM Entrelaçados* (TLE), que se destaca em predições com poucas observações históricas (menos de seis pontos de entrada), superando outras arquiteturas, inclusive a LSTM, nesse escopo de predição.

Em (24), os autores apresentam um modelo *Transformer* aplicado à predição de trajetórias veiculares considerando o cenário de tráfego urbano congestionado, um contexto que também ocorre em grande parte do tráfego para predição de próximo sensor externo, tendo alta complexidade e heterogeneidade. No artigo, a arquitetura *Transformer* utilizada se destaca com redução de cerca de 8,6% de RMSE (*Root Mean Squared Error*) e de 49,9% de MVE (*Maximum Velocity Error*) em relação à LSTM e à MHA-LSTM, que é a LSTM com módulo de atenção multicabeças (*Multi-Head Attention*). Por ser um contexto distinto de previsão, é previsto apenas até 1,2 segundos à frente, bem diferente de uma previsão de próximo sensor externo que chega a ser de horas à frente. Uma visão interessante trazida nesse artigo foi considerar, além

da própria posição, as informações dos veículos ao seu redor, representados em oito direções (Norte, Sul, Leste, Oeste, Nordeste, Sudeste, Noroeste, Sudoeste) e suas interações para prever influências e possíveis colisões em movimentos futuros. Ainda nesse artigo observa-se uma mudança na arquitetura de melhor desempenho, contudo não é abordada essa mudança com uma análise mais detalhada. Outro ponto relevante é que assim como observado em outros artigos e neste trabalho, janelas de entrada mais longas, ainda que neste caso sejam de cerca de 1 segundo, no modelo de predição não necessariamente resultam em predições melhores. Interessante também é o uso do dataset pNEUMA, capturado por drones em interseções de Atenas, na Grécia, que permitiu uma análise diferenciada do cenário urbano, entendido como dados contínuos (25 posições por segundo). Apesar do artigo implementar o MHA-LSTM, essa arquitetura não chega a ser um entrelaçamento de arquiteturas, como a proposta de Transformer e LSTM Entrelaçados (TLE) deste trabalho.

Em (22) os autores utilizam a base de dados contendo sequência de traços MNIST (*MNIST Stroke Sequence*), que são sequências normalmente utilizadas para classificação de números. Contudo, de forma criativa, nesse estudo analisam o trajeto da escrita do número e, com base no trajeto, classificam qual é o número escrito. O interessante é perceber que nessa base de dados são utilizados os dados de fim de sequência parcial, que seria um traço, e o fim da sequência final. Em predição de próximo sensor externo não existem tais informações, porém, em uma das aplicações da nova arquitetura TLE, dividimos as sequências por dia, entendendo de maneira similar, que o fim de um dia seria o fim de uma sequência parcial. Apesar do artigo classificar o trajeto final, nos faz refletir sobre a opção de abordar o problema de próximo sensor externo como de classificação, que faria sentido para um cenário menor e mais controlado, porém para diversos sensores com fluxos de veículos muito diversificados e que podem variar de quantidade e, por vezes, a localização em si, a abordagem como problema de regressão tende a ser mais apropriada. No artigo as sequências chegam a 118 pontos e a arquitetura LSTM superou SimpleRNN e Transformer, que condiz com os resultados obtidos neste trabalho para sequências mais longas.

Os artigos (27), (28), (13) e (19), com alguns autores em comum como CRUZ, L. e MACEDO, J. A. F. e outros individualizados, trazem semelhanças interessantes a este trabalho ao utilizarem bases de dados de sensores externos para predição de trajetória, sendo (27) e (28) abordando registros gerais de veículos comuns e (13) e (19) restringindo para registros de veículos roubados. A seguir analisam-se esses quatro casos um a um em ordem cronológica para um melhor detalhamento.

Em (27), devido à incompletude e esparsidade dos dados dos trajetos veiculares em problemas de *External Sensor Trajectory* (EST), ou trajetória de sensor externo, a ideia principal do artigo é completar os registros dos sensores teoricamente intermediários que o veículo poderia ter passado entre os registros encontrados em uma base de dados reais. Após o preenchimento dos registros intermediários são adotadas duas estratégias para avaliar a predição de próximo sensor externo: uma considerando todos os registros reais e os inseridos avaliando a predição com o registro real original final (*Full Imputation*); e outra alterando o alvo de predição para o próximo registro intermediário inserido após o penúltimo registro real (*Next Value Imputation*) que seria previsto. Não obstante haver a comparação com Modelos de Markov de 1ª e 5ª ordem e adicionalmente considerar estratégia de agrupamento (*Clustering*), nesse artigo também é utilizado apenas um tipo de modelo de IA para predição, nesse caso um modelo de *Recurrent Neural Network* (RNN). Os experimentos indicam que a combinação de estratégias de *Next Value Imputation* com *Clustering* (NIC), passando pelo RNN, atinge as melhores predições, considerando um valor de 5 (para

o *Epsilon* em quilômetros) de agrupamento no algoritmo DBSCAN. Essa combinação de técnicas (NIC) também alcança os menores Erros de Proximidade (*Closeness Error*) para percentis acima de 60%. Uma questão que vale ressaltar são os diversos pontos de transição entre os melhores resultados das variadas combinações de estratégias e parâmetros. Alguns resultados são comentados, porém não há uma análise em relação à alternância de melhor combinação de estratégias. Em contraste, neste trabalho, ocorre a análise de ponto de transição quando há variação de melhor arquitetura de predição.

O artigo (28) é uma continuação adaptada do artigo anterior (27), com a duplicação do modelo RNN e gerando uma nova saída de predição para a dimensão de tempo além da dimensão de espaço que já era gerada no primeiro artigo que passaram a chamar de *single-task* (monotarefa) e o segundo de *multi-task* (multitarefa). Visualmente não fica claro como um novo modelo RNN paralelo prevendo uma dimensão de tempo poderia impactar no resultado do RNN prevendo a posição espacial, todavia, os autores informam que a função de perda utiliza uma média das duas predições para ajustar os pesos dos modelos. Apesar dessa abordagem não ter destaque e não melhorar a predição da maior parte das estratégias do artigo, o impacto dessa nova estratégia *Multi-Task* (MT) é significativo para o conjunto de métodos *Full Imputation* (MT-FI) e *Full Imputation com Clustering* (MT-FIC), chegando a superar em cerca de 20 pontos percentuais os resultados do *Next Value Imputation com Clustering* (NIC). Uma parte dos gráficos do artigo mostram informações incoerentes com os resultados descritos, contudo de forma geral é observável um ganho com a nova abordagem. Essa nova estratégia multitarefa não chega a ser uma nova arquitetura de IA para predição, mantendo duas RNN, tampouco um entrelaçamento de arquiteturas, contudo demonstra que combinações de informações como o tempo podem aprimorar a predição de próximo sensor externo, como acontece neste trabalho.

Já em (13) os autores trazem a ideia de incluir atributos semânticos como fatores de auxílio à predição. Duas categorias são percorridas: atributos criminais, considerando informações de quantidade e tipos de crimes; e atributos de pontos de interesse (*Point of Interest*), que seriam locais onde os criminosos costumariam abandonar os veículos roubados. Percebe-se que os autores consideram atributos semânticos extrínsecos ao veículo que se pretende prever a próxima geolocalização, em contraponto, neste trabalho é apresentado um atributo adicional intrínseco, que é a via da rua que o veículo está ao passar por um sensor externo. O artigo retrata quatro métodos de predição: *K-Nearest Neighbor* (KNN), *Random Forest* (RF), *Gradient Boost Regression Tree* (GBRT) e *Bidirectional LSTM* (BiLSTM). Embora o artigo realize comparações e um pré-processamento interessantes, ele não traz de fato uma nova arquitetura de IA como acontece neste trabalho. Todavia, no artigo a arquitetura GBRT, em uma configuração chamada em outros contextos de *eXtreme Gradient Boosting* (XGBoost), supera as demais e, é notável por ser o único trabalho nesse conjunto a utilizar o XGBoost, destacando seu potencial para lidar com dados esparsos e discretos, particularmente na predição da localização de veículos roubados, obtendo os melhores resultados nesse artigo. Em razão disso, essa arquitetura foi incorporada a alguns experimentos deste trabalho. Outro ponto de ênfase desse artigo é a análise de representação espacial, onde compara três representações: baseado em Geohash, onde geo-hash é uma função que codifica coordenadas em um único valor; baseado em Grade (*Grid*), onde divide-se virtualmente o espaço em células indicadas por linhas e colunas; e baseado em *Cluster*, onde o espaço é dividido em conjuntos aplicando-se o algoritmo DBSCAN. Nesse artigo, apesar de ser utilizado um valor de precisão cerca de dez vezes acima do que o método Geohash teoricamente necessitaria, esse método obteve o melhor resultado. O artigo fixa em cinco a quantidade de pontos de

sequência histórica de entrada nos modelos de IA, o que permite observar variação na predição ao variar a forma de representação espacial. No entanto, não permite avaliar de maneira mais abrangente o impacto da variação da sequência histórica combinada com diferentes representações geográficas, como é apresentado neste trabalho.

Em (19) os autores adotam uma camada de Processamento de Linguagem Natural - PLN, valendo-se da capacidade de associação e relação entre as palavras, para que, de forma similar, essa camada incorpore e realize a representação espacial de um mapa geográfico traduzindo diferentes trajetórias veiculares de sensores externos. Nessa camada são comparados os resultados de cinco formas diferentes de incorporação (*embedding*) seguidas da arquitetura LSTM: uma camada de *embedding* treinada do zero chamando o todo apenas de LSTM; uma que explora um *embedding* pré-treinado BERT MLM, chamando o todo de LBERT; uma que vai além e realiza ajuste fino (*Fine-Tuned*) do BERT MLM, chamando de LBERT-FT; uma que emprega Word2Vec, chamada de LW2V; e uma onde dispõe do ajuste fino (*Fine-Tuned*) do Word2Vec, chamando de LW2V-FT. Nesse artigo o LBERT-FT atinge os melhores resultados, mostrando que essa camada que antecipa o modelo LSTM e reduz a dimensionalidade ao mesmo tempo em que incorpora informações de relacionamento entre os pontos, favorece a predição do próximo sensor externo. Neste trabalho utiliza-se apenas uma camada densa (*Fully-Connected*) para realizar essa atividade de *embedding* na entrada dos modelos de IA quando necessário, com o intuito de generalização e independência de modelos pré-treinados externos. Interessante observar que no artigo são apresentados resultados de diferentes tamanhos de sequência histórica (5, 7, 15 e 31 pontos) e que, quanto mais pontos históricos, piores são os resultados. Esses resultados estão em consonância com uma parte deste trabalho para grandes quantidades de pontos históricos, quando 30 pontos atingem uma predição inferior a 10 pontos na sequência de entrada do modelo de predição. conquanto seja uma análise enriquecedora, o artigo não emprega um novo modelo ou diferentes modelos de IA após a camada de *embedding*.

3.2 NOVA ARQUITETURA DE PREDIÇÃO

Em (15), apesar de não ter criado uma arquitetura nova, trata-se de um dos primeiros estudos a utilizar a arquitetura Transformer para predição de destino sem informações adicionais de contexto, ou seja, utilizando apenas as geolocalizações como entrada no modelo de predição. O artigo conduz experimentos com duas bases de dados: uma simulada gerada com o *software* SimTread, que cria trajetórias de movimento em ambientes internos, e outra, mais próxima da realidade deste trabalho, que é um conjunto de dados real, composto por trajetórias de táxis na cidade de Porto, em Portugal. Para esse artigo, a sequência de entrada consiste na trajetória contínua parcial e a saída é o ponto final da trajetória, variando o tamanho do trecho ausente entre a trajetória inicial parcial e o ponto de destino. Os resultados mostram uma certa discrepância entre os experimentos com dados simulados, onde a arquitetura Transformer apresentou uma superioridade bem significativa e os experimentos com dados reais, onde os gráficos demonstram similaridade com os resultados deste trabalho, onde há uma transição de melhor arquitetura e a arquitetura LSTM atinge melhores predições ao aumentar a quantidade de dados de entrada para a predição.

Em (25), é abordado o desafio de predição de trajetória envolvendo veículos autônomos e demais veículos à sua volta durante os trajetos. Os autores propõem um modelo *intention-aware Non-Autoregressive*

Transformer com aprendizado *multi-attention* (iNATran) para a predição de trajetórias veiculares. Esse modelo combina módulos dedicados de aprendizado social e temporal: um módulo de atenção social, que utiliza grafos para capturar relações entre veículos próximos; e um módulo de atenção temporal, responsável por modelar as dependências entre quadros consecutivos. Além disso, um decodificador implementa a predição multimodal, gerando múltiplas trajetórias simultaneamente e reduzindo o tempo de inferência, o que gera eficiência computacional no contexto do *Internet of Vehicles* (IoV). Essas inovações mostram ganhos variando de 5,9% a 9,3% em relação às arquiteturas mais comuns utilizadas anteriormente nas bases de dados NGSIM e HighD utilizadas. Neste trabalho a nova arquitetura também é baseada na estrutura original da arquitetura *Transformer*, bem como altera a arquitetura internamente e não só antes ou após, como na maioria dos outros estudos. O artigo propõe módulos específicos para as necessidades do contexto, porém não utiliza uma arquitetura entrelaçada, como acontece neste trabalho. Outro ponto a observar é que a predição é realizada a no máximo cinco segundos à frente da posição atual do veículo e com base em três segundos anteriores, o que é bem diferente da realidade de predição de próximo sensor externo onde essa latência entre pontos pode ser de horas por vezes.

Em (16) uma visão de combinação de grafos é empregada de forma bem interessante, tendo um grafo local dinâmico construído com base em um grafo global, onde o grafo global modela a conectividade de toda a área de estudo, e os grafos locais representam subestruturas dinâmicas com base em trajetórias individuais e pontos de interesse (*Point of Interest* - POI). Para modelos de predição, são aplicadas variações do estilo LSTM (LSTM, GRU, BiLSTM, Attn-LSTM e Attn-BiLSTM), e a proposta dos autores de uma nova arquitetura chamada de LDGST-LSTM (*Local Dynamic Graph Spatiotemporal-Long Short-Term Memory*). Para os experimentos foram utilizados dados reais de táxis da cidade de Chengdu, na China. O ganho da arquitetura LDGST-LSTM proposta foi significativo, atingindo um percentual médio de 3,5 a 4,5 vezes superior em relação às demais arquiteturas, dependendo do cenário avaliado. Interessante notar que os ganhos foram mais expressivos em cenários com maior complexidade espacial, indicando que, por vezes, uma arquitetura mais complexa pode ser capaz de capturar melhor a predição de cenários proporcionalmente mais complexos. De maneira similar, a arquitetura *Transformer* e LSTM Entrelaçados (TLE) é capaz de superar as predições individuais das arquiteturas *Transformer* e LSTM para determinados cenários, especialmente para sequências históricas mais curtas.

Em (26), uma publicação de nossa autoria, são apresentados resultados que demonstram que ao incluir as informações sobre qual via da rua o veículo foi identificado ao passar por um sensor externo, como um radar de velocidade, a predição de próximo sensor externo é melhorada em média 6,5%, destacando a importância de examinar atributos intrínsecos ao veículo observado. Essa estratégia também é utilizada nas aplicações presentes neste trabalho. A arquitetura *Transformer* e LSTM Entrelaçados (TLE) foi apresentada pela primeira vez nesse artigo, em comparação às arquiteturas LSTM e *Transformer*, podendo ser observado um potencial de ganho de predição da TLE em cenários de poucos pontos na sequência histórica de entrada do modelo de IA para a predição de próximo sensor externo e a LSTM se destacando para sequências de entrada maiores, como vimos ocorrer de maneira similar em outros estudos. Esse estudo observa uma mudança na arquitetura de melhor desempenho, mas carece de uma análise detalhada do ponto de transição entre essas arquiteturas. Interessante notar que, apesar dos treinamentos serem de certa forma demorados para modelos de *deep learning* complexos, não sendo viável o treinamento completo ser dinâmico, a predição em si é rápida e, por isso, só é realizada a predição de próximo sensor externo,

em vez de vários de uma vez. Assim, pode ser realizada nova predição de próximo sensor externo a cada novo sensor que o veículo seja identificado. Como visto em outros estudos, quanto mais longe o ponto da sequência de predição do último ponto real identificado, menos precisa a predição é. Considerando esse contexto veicular, neste trabalho também é aplicada a predição de próximo sensor externo.

No contexto do presente trabalho, que considera XGBoost, LSTM, Transformer e a arquitetura TLE, abordamos as lacunas e a inclusão da informação de qual via o veículo é identificado pelo sensor externo, além de examinar detalhadamente a mudança na arquitetura de melhor desempenho, conduzindo uma análise do ponto de transição e explorando o impacto da variação na representação geográfica, observando em destaque a arquitetura TLE. Essa abordagem abrangente visa melhorar a predição da trajetória do próximo sensor externo, especialmente no desafiador contexto de dados esparsos e discretos associados a veículos roubados. O resumo comparativo é apresentado na **Tabela 3.1**.

Tabela 3.1: Resumo comparativo entre estudos relacionados.

Artigo	Informação Adicional para Predição	Arquiteturas de Predição	Particularidades
(23)	Coordenada Z (altura) para trajetórias balísticas.	LSTM, GRU, MLP e Transformer.	Compara várias arquiteturas para prever trajetórias de veículos balísticos em 3D.
(24)	Interações sociais entre veículos.	LSTM, MHA-LSTM e Transformer.	Considera interações entre veículos em cenários urbanos congestionados, prevendo posições futuras para evitar colisões.
(22)	Fim de traço e fim de sequência na escrita manuscrita.	LSTM, SimpleRNN e Transformer.	Aborda a classificação de números manuscritos com base em trajetórias de escrita.
(27)	Inserção de registros intermediários.	RNN.	Trabalha com dados esparsos de sensores externos e aplica estratégias de inserção de valores ausentes.

Continuação da Tabela 3.1: Resumo comparativo entre estudos relacionados.			
Artigo	Informação Adicional para Predição	Arquiteturas de Predição	Particularidades
(28)	Dimensão temporal simultânea.	Markov e RNN.	Introduz multitarefa na predição, considerando predições de tempo e espaço simultaneamente.
(13)	Atributos criminais e pontos de interesse (POIs).	KNN, RF, GBRT (XGBoost) e BiLSTM.	Utiliza atributos extrínsecos ao veículo, como dados criminais e POIs, destacando a arquitetura GBRT (XGBoost).
(19)	Relações de Processamento de Linguagem Natural (PLN) para representar mapas.	LSTM, LBERT, LBERT-FT, LW2V e LW2V-FT.	Utiliza embeddings pré-treinados de PLN (BERT MLM e Word2Vec) para melhorar a representação espacial em predições veiculares.
(15)	Nenhuma (apenas dados de geolocalização).	LSTM, BiLSTM e Transformer.	Um dos primeiros a aplicar Transformer para predição de destinos veiculares sem contexto adicional, usando dados de táxis.
(25)	Intenções dos veículos e interações sociais.	S-LSTM, CS-LSTM, S-GAN, HMNet, MHA-LSTM, STDAN e iNATran.	Propõe uma predição multimodal para trajetórias de veículos no contexto de veículos autônomos, incluindo módulos sociais e temporais.

Continuação da Tabela 3.1: Resumo comparativo entre estudos relacionados.			
Artigo	Informação Adicional para Predição	Arquiteturas de Predição	Particularidades
(16)	Pontos de Interesse (POIs) e grafos dinâmicos locais.	LSTM, GRU, BiLSTM, Attn-LSTM, Attn-BiLSTM e LDGST-LSTM.	Combina grafos locais dinâmicos e globais para aprimorar predições de trajetórias em cenários complexos.
(26)	Informação da via onde o veículo foi detectado.	LSTM, Transformer e TLE.	Introduz a arquitetura Transformer e LSTM Entrelaçados (TLE), destacando a relevância de atributos intrínsecos.
Este Trabalho	Informação da via onde o veículo foi detectado e variação do formato de representação geográfica.	XGBoost, LSTM, Transformer e TLE.	Arquitetura TLE aplicada usando informação da via da rua e variação do formato de representação geográfica, analisando transição entre melhores arquiteturas. Único com informação da via. Único com TLE. Único com análise do ponto de transição entre arquiteturas. Único com análise de impacto da variação da representação geográfica.

3.3 CONSIDERAÇÕES FINAIS

Nos artigos analisados, pode-se observar que diversas abordagens são aplicadas ao contexto de predição de trajetória veicular e que utilizar dados esparsos gera uma complexidade maior que para dados contínuos. Incluir novas informações costumam aprimorar a predição de geolocalização veicular, sendo uma abordagem interessante a se considerar nesse contexto. Outro aspecto relevante é que a arquitetura LSTM é uma das principais utilizadas e que tende a ser vantajosa em sequências históricas maiores como entrada no modelo de IA para predição. Arquiteturas mais complexas, por vezes, se destacam em relação ao LSTM em sequências de entrada menores, como é o caso da arquitetura Transformer e LSTM Entrelaçados (TLE). No Capítulo 4, apresentamos a metodologia utilizada neste trabalho e, na sequência, no Capítulo 5 detalhamos a construção da arquitetura TLE.

4 METODOLOGIA E ARQUITETURA ENTRELAÇADA

Neste Capítulo será apresentada a metodologia, seguida ao longo da jornada de construção do trabalho dessa dissertação, para atingir os resultados alcançados e os fundamentos de como ocorreu a concepção da arquitetura entrelaçada (Transformer e LSTM Entrelaçados - TLE), assim como os detalhes envolvidos em cada fase.

4.1 FLUXOGRAMA DA METODOLOGIA

Um trabalho desse porte naturalmente passa por diversas etapas e, se não nos organizarmos, é fácil se perder em meio a tantas informações. Pensando nisso, na Figura 4.1 é apresentado um fluxograma do desenvolvimento deste trabalho. Importante esclarecer que essa é a metodologia do trabalho da dissertação, havendo, no entanto, dois estudos mais específicos realizados e que, naturalmente, houve algumas nuances em cada um, que são detalhadas neste Capítulo e nos Capítulos 6 e 7.

4.1.1 Definição do Trabalho

Começando pela definição do trabalho, o problema, provavelmente também percebido por outras unidades da federação, foi identificado e comentado por servidores públicos do Ministério da Justiça e Segurança Pública (MJSP), que viam constantemente grandes quantidades anuais de roubos de veículos e perceberam que esse é um crime que costuma ser predecessor de outros delitos. Destarte, em sentido lógico, ao recuperar rapidamente um veículo roubado é interrompida uma cadeia de outras transgressões. Nesse contexto, sabe-se que os dados de geolocalizações, geralmente, são valores contínuos. Isso significa que, ao prever a posição de um veículo para auxiliar sua recuperação, o mais adequado é prever a possibilidade de valores também contínuos. Dessa maneira, determinou-se que essa situação seria tratada como um problema de

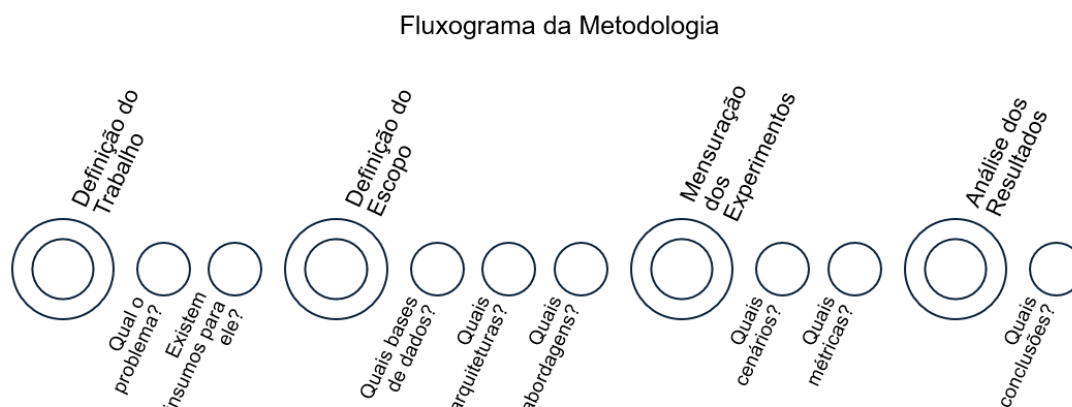


Figura 4.1: Fluxograma da Metodologia.

regressão, haja vista os valores de latitude e longitude serem contínuos tornando o problema de regressão a forma mais adequada de tratar esse contexto (43). Destacamos essa definição da **visão geral do problema** a seguir:

- Visão geral do problema: Problema de regressão.

A partir disso, vem o questionamento se existiriam insumos - informações e recursos - que apoiariam uma busca mais eficiente pela recuperação desses bens. Em busca por esses insumos, notou-se que raramente um veículo roubado possui um *Global Positioning System* (GPS) transmitindo dados em tempo real ou mesmo posteriormente. Entretanto, descobriu-se que é possível identificar a localização de veículos roubados, mesmo que de forma pontual e esparsa, por meio de câmeras como em radares de velocidade. Em vista disso, fez-se uma reflexão em relação a abordagem como um problema de regressão, porquanto os dados obtidos serem esparsos e pontuais. Poderia-se pensar em uma abordagem de classificação, porém seria difícil generalizar o modelo de IA treinado, como a possibilidade de prever pontos que não apareceram nos dados de treinamento. Outrossim, em relação aos sensores externos, apesar de quando instalados terem uma posição fixa, por vezes, eles são movidos de posição, bem como são instalados novos e removidos outros. Essas são situações normalmente adversas a problemas de classificação e, em caso de dados desbalanceados, um problema de classificação seria mais prejudicado que um problema de regressão. Após tal reflexão, ratificamos a adoção como um problema de regressão e que existem **insumos**, como destacado a seguir:

- Insumo: registros dos veículos ao passar por radares de velocidade.

4.1.2 Definição do Escopo

Identificado o problema e que existem insumos plausíveis para abordá-lo, precisa-se detalhar melhor o escopo do que seria utilizado e com quais objetivos específicos. Nesse sentido, precisou-se identificar quais bases de dados e delas, quais dados seriam consumidos para este trabalho. De imediato ficou evidente a necessidade das localizações dos veículos com carimbo de data e horário (*timestamp*). No entanto, foram necessárias duas bases para isso, uma com os registros de passagem dos veículos pelos sensores contendo um código identificador do veículo, o carimbo de tempo e um código identificador da câmera (sensor externo, como um radar de velocidade) que realizou a identificação. A outra base de dados contém o código identificador do sensor externo e a localização em latitude e longitude. Logo, executando um cruzamento entre essas bases é possível chegar na geolocalização pontual de um veículo. Nesse ponto surge outra questão, que é saber quais de todos os veículos que passam por esses radares são roubados. Para tal, uma terceira base de dados foi necessária contendo os registros de boletins de ocorrências contendo quais veículos estão com ocorrência de roubo ativa.

Por serem bases de dados com grande volume de dados, alguns cruzamentos e recortes foram aplicados para que fosse viável trabalhar com os dados. Primeiro, cruzou-se a base dos registros de passagens em radares com os registros de boletins de ocorrências, assim gerando uma base única de alertas, contando apenas registros de passagens apenas dos veículos com boletins de ocorrência. Mesmo reduzindo significativamente o volume de registros e de três para duas bases de dados, ainda assim eram muitos dados para se

trabalhar. Para se ter uma ideia, a quantidade de registros armazenados em uma base de dados SQL Server na nuvem Azure estava atingindo o limite de armazenamento e processamento dos dados, mesmo considerando a escalabilidade da nuvem. Portanto, mesmo uma parcela de todo esse volume ainda era muito para se trabalhar com treinamento de modelos de *deep learning*. Por essa razão, para o primeiro estudo, objetivando “Ganho de Predição”, fez-se necessário uma filtragem por três Estados do Brasil, de acordo com os seguintes critérios: Ceará (CE) por conter dados relacionados a outros estudos sobre a mesma temática; Distrito Federal (DF) por ser onde moro e saber que possui uma boa quantidade de radares e Goiás (GO) por ir com certa frequência de carro para Goiânia. Assim, haveria informações além das bases de dados recebidas, caso fosse necessário algum tipo de validação extra. Então, para o segundo estudo, objetivando “Apoio à Definição de Melhor Arquitetura”, optou-se por um único Estado, porém com o maior volume de registros, determinando, então, a base de dados do Estado de São Paulo (SP). Logo, as **bases de dados** elencadas para os respectivos estudos foram:

- Ceará (CE): 1º estudo;
- Distrito Federal (DF): 1º estudo;
- Goiás (GO): 1º estudo; e
- São Paulo (SP): 2º estudo.

Além de delimitar quais dados utilizar nos estudos, fundamental também é definir quais arquiteturas de aprendizado de máquina seriam implementadas para os treinamentos e predições. Após, pesquisas de artigos sobre os assuntos de predição de trajetória e rota veicular, tanto de sensor externo quanto de dados contínuos, identificou-se que a arquitetura *Long Short-Term Memory* (LSTM) era a mais utilizada e com destaque em relação a outras alternativas. Da mesma forma, a arquitetura Transformer também apresentava destaque em artigos, havendo um certo equilíbrio e alternância entre LSTM e Transformer em relação a qual era a melhor arquitetura em cada estudo. Diante do exposto, inicialmente foram elencadas LSTM e Transformer para o primeiro estudo, objetivando “Ganho de Predição”. No entanto, para se implementar uma arquitetura de IA é essencial entender como ela funciona. Ao compreender o funcionamento da LSTM e da Transformer e implementá-las aplicando testes iniciais de funcionamento, percebeu-se que essas arquiteturas de *deep learning* funcionam com um nível de aprendizado tão profundo e atingindo resultados tão impressionantes que os seres humanos têm dificuldades de explicar exatamente como isso ocorre internamente. Entretanto, uma característica em comum é que boa parte das arquiteturas são de camadas de neurônios (simulação digital do funcionamento de um neurônio humano). Assim, se valendo desses entendimentos, realizei diversos testes e criei uma nova arquitetura, entrelaçando as duas principais sobre o tema abordado, assim nasceu a arquitetura Transformer e LSTM Entrelaçados (TLE) e foi incorporada na lista de arquiteturas do primeiro estudo, para colocar à prova essa inovação. Para o segundo estudo, objetivando “Apoio à Definição de Melhor Arquitetura”, buscamos uma arquitetura diferente que também tivesse mostrado grande potencial em algum estudo e encontramos a *Extreme Gradient Boosting* (XGBoost), que superou a LSTM em (13). Dessa maneira, as **arquiteturas** empregadas neste trabalho e seus respectivos estudos foram:

- LSTM: 1º e 2º estudos;

- Transformer: 1º estudo;
- Transformer e LSTM Entrelaçados (TLE): 1º e 2º estudo; e
- XGBoost: 2º estudo.

O terceiro ponto importante para definição do escopo é quais abordagens serão empregadas nos experimentos. A nova arquitetura TLE já era uma novidade por si só, porém ela deveria ser utilizada como mais uma arquitetura para validar um outro conceito e, assim, mostrar suas capacidades de predição de forma comparativa e contextualizada, em um rol de concorrência de estado-da-arte. Por conseguinte, analisando os dados disponíveis, para o primeiro estudo identificou-se no melhor do entendimento dos autores, que utilizar a informação de via da rua que o veículo foi identificado para auxiliar na predição, até então, nunca se tinha ocorrido, sendo a primeira abordagem a ser executada.

Com os resultados do primeiro estudo, identificou-se uma transição entre a melhor arquitetura a ser empregada a depender do tamanho da sequência histórica de entrada no modelo de IA. Com essa observação, um novo estudo foi planejado para aprofundar o entendimento dessa situação, sendo a segunda abordagem identificar o provável ponto de transição entre as melhores arquiteturas, agora em uma outra base de dados, reduzindo possíveis vieses das já utilizadas. A terceira abordagem foi variar o formato de representação geográfica das geolocalizações analisadas a fim de observar os impactos dessa variação tanto nas predições, quanto no ponto de transição entre arquiteturas. Assim, as **abordagens gerais** analisadas, de forma a comparar a arquitetura TLE com outras alternativas de arquiteturas no estado-da-arte, e seus respectivos estudos foram:

- Inclusão da informação sobre qual via da rua o veículo foi identificado: 1º estudo;
- Identificação do ponto de transição entre as melhores arquiteturas de predição de próximo sensor externo: 2º estudo; e
- Identificação do impacto da variação do formato de representação geográfica das geolocalizações nas predições e no ponto de transição entre as melhores arquiteturas de predição de próximo sensor externo: 2º estudo.

4.1.3 Mensuração dos Experimentos

Para se comparar performances de diferentes experimentos e arquiteturas é essencial aferir os resultados com métricas que se apliquem ao contexto dos cenários executados. Assim sendo, devemos definir quais cenários serão executados.

A abordagem geral de *“inclusão da informação sobre qual via da rua o veículo foi identificado”* foi implementada para o primeiro estudo, objetivando “Ganho de Predição”, onde os modelos de IA LSTM, Transformer e TLE foram treinados com os registros dos estados do Ceará (CE), Distrito Federal (DF) e Goiás (GO) considerando os dados sobre qual via da rua o veículo foi identificado e desconsiderando esse dado, para avaliar se essa informação melhora a predição ou se, ao contrário, gera ruído no treinamento dificultando a predição.

Em continuidade, as abordagens de “*identificação do ponto de transição entre as melhores arquiteturas de predição de próximo sensor externo*” e de “*identificação do impacto da variação do formato de representação geográfica das geolocalizações nas predições e no ponto de transição entre as melhores arquiteturas de predição de próximo sensor externo*” foram implementadas para o segundo estudo, objetivando “Apoio à Definição de Melhor Arquitetura”. Nesse contexto, os modelos de IA XGBoost, LSTM e TLE foram treinados com os registros do estado de São Paulo (SP), variando o tamanho da sequência histórica de entrada de 1 a 8 geolocalizações para encontrar o ponto de transição entre as arquiteturas com melhores resultados e comparando os formatos de representação geográfica EPSG:4326 (WGS84), EPSG:3857 (Web Mercator) e EPSG:32723 (UTM).

Tendo em mente que é um problema de regressão e os cenários já identificados, importante definir quais métricas usar em cada um para uma melhor comparação dos resultados. Sob tal perspectiva, como métricas de avaliação para o primeiro estudo, foram utilizadas as métricas Raiz do Erro Médio Quadrático (*Root Mean Squared Error* - RMSE), Erro Médio Quadrático (*Mean Squared Error* - MSE) e Erro Médio Absoluto (*Mean Absolute Error* - MAE) que são métricas convencionais para problemas de regressão (44) (45), onde todas essas três métricas se comportam de forma que quanto menor o seu valor melhor o desempenho do modelo. Ainda que RMSE e MSE tenham uma relação direta matemática, são apresentadas as duas informações para auxiliar possíveis análises comparativas posteriores em outros estudos.

Para avaliar os resultados do segundo estudo, as métricas utilizadas foram o Erro Médio Quadrático (*Mean Squared Error* - MSE), mantendo uma métrica que possa manter uma correlação mais direta com o primeiro estudo, e o Coeficiente de Determinação (R^2) como uma nova métrica para ampliar as métricas de avaliação das arquiteturas analisadas. O MSE é uma métrica padrão para problemas de regressão (44) (45), quantificando a média dos erros quadráticos entre valores previstos e reais, enquanto o R^2 , ou coeficiente de determinação, é comumente usado em *machine learning* como uma métrica de desempenho para modelos de regressão, indicando a proporção da variabilidade dos dados que é explicada pelo modelo. Ele ajuda a determinar o quão bem um modelo se ajusta aos dados. O valor de R^2 varia de 0 a 1, onde 0 indica que o modelo não explica nenhuma variação, e 1 indica que o modelo explica toda a variação observada (46) (47). Portanto, para o MSE, valores menores indicam melhor desempenho, e para o R^2 , valores maiores indicam melhor desempenho. Assim sendo, as **métricas avaliativas** para os respectivos estudos foram:

- RMSE, MSE e MAE: 1º estudo; e
- MSE e R^2 : 2º estudo.

4.1.4 Análise dos Resultados

Para que os resultados fossem robustos os experimentos foram extensos. No primeiro estudo, objetivando “Ganho de Predição”, ocorreram 18 cenários de treinamentos para cada uma das três arquiteturas (LSTM, Transformer e TLE), totalizando 54 cenários, sendo cada conjunto de treinamento e predição executado três vezes e considerado apenas o melhor valor dos três para as análises. Isto significa um total de 162 execuções de treinamentos e mais 162 execuções de predições para esse contexto.

Para o segundo estudo, objetivando “Apoio à Definição de Melhor Arquitetura”, algumas mudanças foram efetuadas para ampliar a robustez dos resultados. Em vez de executar três vezes o mesmo conjunto

de dados e considerar o melhor valor, as avaliações foram realizadas duas vezes para cada conjunto de cenários (arquitetura, tamanho de sequência e representação geográfica): primeiro utilizando os últimos 20% dos dados como conjunto de teste e, depois, os primeiros 20%, adotando o valor final da métrica como a média dos resultados obtidos em cada teste. Essa abordagem é semelhante a uma validação cruzada k-fold parcial, garantindo uma análise robusta, porém eficiente. Nesse segundo estudo foram 14 cenários de treinamentos para cada arquitetura, totalizando 42 cenários, sendo cada conjunto de treinamento e predição executado duas vezes (com parcelas de dados distintas) e considerando apenas a média. Com essa atuação realizou-se um total de 84 execuções de treinamentos e mais 84 execuções de predições para esse contexto. Portanto, o total de **execuções** realizadas nos respectivos estudos foram:

- 1º estudo: 54 cenários, 162 execuções de treinamentos e mais 162 execuções de predições; e
- 2º estudo: 42 cenários, 84 execuções de treinamentos e mais 84 execuções de predições.

O detalhamento dos experimentos, assim como as análises dos resultados e suas conclusões são apresentados, respectivamente, no Capítulo 6 para o primeiro estudo e no Capítulo 7 para o segundo estudo. Para implementar os experimentos, processar os dados e plotar em mapas os pontos e trajetos das figuras apresentadas neste trabalho, foram utilizados a linguagem *python*, as bibliotecas *TensorFlow* e *Folium* e o *OpenStreetMap* na plataforma *JupyterLab*, executado em um notebook com um processador Intel i7 de 12ª geração, 16GB de RAM e um SSD de 1TB.

4.2 CONSIDERAÇÕES FINAIS

O Fluxograma da Metodologia apresentou uma visão sistêmica de como ocorreu a definição do problema abordado neste trabalho. As nuances do Ministério da Justiça e Segurança Pública (MJSP) e a intenção de interromper a cadeia de crimes iniciada por um roubo de veículo são admiráveis. Tanto como a definição do escopo das quatro bases de dados (CE, DF, GO e SP) estudadas. A escolha por arquiteturas em seu estado-da-arte (XGBoost, LSTM e Transformer) com destaque em outros artigos para comparar com a nova arquitetura TLE demonstra a robustez dos resultados deste trabalho. Além da robustez, a versatilidade fica por conta das abordagens analisadas, incluindo dados sobre qual via da rua o veículo foi identificado, identificação do ponto de transição entre as melhores arquiteturas e a identificação do impacto da variação do formato de representação geográfica nas predições e no ponto de transição entre as melhores arquiteturas de predição de próximo sensor externo. Também vimos que foram empregadas métricas adequadas para a mensuração dos experimentos (RMSE, MSE, MAE e R^2), que atingiram um total de 246 execuções de treinamentos e mais 246 execuções de predições.

5 ARQUITETURA ENTRELAÇADA TLE

Quando o *generative pre-trained transformer 3* (GPT-3) desbancou o LSTM na área de Processamento de Linguagem Natural (PLN), gerou-se a expectativa de que a arquitetura Transformer seria melhor do que a LSTM em previsões de sequências, mesmo em áreas distintas de PLN. No entanto, existem artigos nos quais um modelo LSTM tem melhor performance do que modelos Transformer, como em (22) e (23), e em outros artigos o modelo Transformer obtém melhores resultados como em (25) e (15). Logo, ainda não se tem uma arquitetura com desempenho superior na área de predição de trajetória veicular independente do contexto em que é inserida.

Em (19) há o uso de uma arquitetura híbrida de LSTM e Transformer, utilizando o modelo pré-treinado BERT (48). Naquele contexto, apesar de sua característica híbrida ser sequencial, essa arquitetura híbrida obteve os melhores resultados, deixando um viés de superação para uma arquitetura híbrida em relação às arquiteturas LSTM e Transformer individualizadas.

O mecanismo de aprendizado por autoatenção (*Self-Attention Learning*), como utilizado na arquitetura Transformer (3) (26), permite que o modelo capture dependências globais ao longo de toda a sequência sem a necessidade de processar os dados em ordem sequencial. Esse mecanismo possibilita que o modelo tenha “atenção” a diferentes partes da sequência simultaneamente, capturando relações de longo alcance e padrões complexos que podem ocorrer em qualquer ponto da série de dados. Por outro lado, o aprendizado sequencial, representado pelas LSTMs, é especializado na modelagem de dependências temporais ao processar os dados passo a passo. A LSTM armazena informações ao longo do tempo, permitindo que o modelo aprenda padrões dependentes da ordem dos eventos e mantenha um histórico por meio de suas células de memória interna.

Ao entrelaçar esses dois mecanismos na arquitetura Transformer e LSTM Entrelaçados (TLE), o modelo tem o potencial de aproveitar os pontos fortes de ambas as abordagens. O Transformer captura dependências globais dentro da sequência, enquanto a LSTM adiciona uma camada de aprendizado temporal que preserva a ordem dos eventos. Essa combinação permite que a TLE aprenda tanto padrões globais e relações estruturais quanto dependências temporais detalhadas, resultando em um modelo capaz de capturar nuances mais complexas em dados sequenciais.

Criar uma arquitetura de *deep learning* não é uma tarefa trivial. Por conseguinte, criar uma arquitetura de *deep learning* que supere o estado-da-arte de outras arquiteturas já criadas, testadas e validadas por diversas pessoas e estudos é um desafio ainda maior e de inovação, pois requer além do domínio técnico aprofundado, a criatividade para inovar com soluções originais que superem as performances dos padrões estabelecidos pela pesquisa já produzida.

Conforme definido pelo Dicionário Priberam da Língua Portuguesa, inovar consiste em “introduzir novidades ou mudanças.”. Durante os estudos para entendimento e construção das arquiteturas LSTM e Transformer, percebeu-se uma oportunidade de inovação. A arquitetura Transformer possui uma construção do todo mais complexa que a LSTM, contudo, ambas apresentam camadas de redes neurais profundas. Em outras áreas de arquiteturas de *deep learning* como para visão computacional é comum haver variadas

combinações de camadas (*layers*) com técnicas distintas. Outrossim, *deep learning* agrega características de experimentação e que a forma como um modelo aprende, por vezes, não é totalmente clara e explicável, reforçando a questão da experimentação. Baseado nessas informações, iniciou-se um processo de experimentação de entrelaçar essas duas principais arquiteturas - LSTM e Transformer -, inserindo camadas (*layers*) de LSTM em pontos estratégicos da arquitetura Transformer. A cada inserção executavam-se alguns testes para verificar possível ganho de performance, uns pontos demonstrando melhoria e outros o oposto. Assim, após muitas experimentações e testes, chegou-se a um estado-da-arte dessa nova arquitetura LSTM e Transformer Entrelaçados (TLE).

Dedicar-se a somar as vantagens dessas duas arquiteturas LSTM e Transformer já foi visto em outros estudos como a arquitetura MHA-LSTM em (25) e (24), e Attn-LSTM e Attn-BiLSTM em (16), entre outras, que combinam partes das arquiteturas de forma serial, ou seja, uma parte LSTM seguida de uma parte Transformer ou o inverso, ainda assim, sem evidenciar superioridade nos respectivos estudos. No melhor do entendimento dos autores, não houve, até então, a aplicação dessas duas arquiteturas entrelaçadas, especialmente da forma como definida neste trabalho e aplicada em (26).

A Figura 5.1 representa o funcionamento interno da arquitetura Transformer e LSTM Entrelaçados (TLE), apresentando na cor azul, com um asterisco embaixo, as camadas inspiradas na arquitetura LSTM e na cor verde, sem asterisco, as camadas inspiradas na arquitetura Transformer. Como pode ser observado, existem três principais conjuntos de funcionalidades e redes neurais:

- Vetorização (*Embedding*): responsável por projetar e redimensionar as características dos dados de entrada em uma dimensão fixa de 64 variáveis contínuas, ampliando a capacidade do modelo de extrair padrões e representar relações.
- Aprendizado (*Learning*): responsável por aprender as nuances das sequências, tanto em aspectos de longo prazo quanto de curto prazo.
- Predição (*Predicting*): responsável por utilizar o aprendizado da camada anterior para gerar a predição de próxima geolocalização.

O formato dos dados de entrada desse modelo de *deep learning* TLE tem a forma comum para modelos de aprendizado sequencial: tamanho do lote (batch size); comprimento da sequência (time steps); e quantidade de atributos (number of features). Tendo os dados no formato correto, a entrada da sequência histórica de pontos de geolocalização se dá pela camada de Vetorização (*Embedding*), formada por uma camada densa pura, ou seja, sem tratar de “tokens” de palavras, como ocorre em Processamento de Linguagem Natural. Essa camada realiza uma projeção expandida dos dados de entrada - contendo os atributos (*Features*) para treinamento ou predição - em um espaço de maior dimensão e tamanho fixo (64 variáveis contínuas). Isso permite destacar características dos dados como se estivéssemos “dando um zoom” em partes importantes deles, mesmo que nessa analogia seja sem aumentar a resolução original ainda é possível extrair melhores informações.

A camada de Vetorização (*Embedding*) então entrega os dados em espaço expandido para a estrutura de Aprendizado (*Learning*). Nessa parte temos quatro Blocos Transformer (*Transform Blocks*) em série seguidos de uma camada LSTM e uma camada densa. Internamente, cada Bloco Transformer (*Transform*

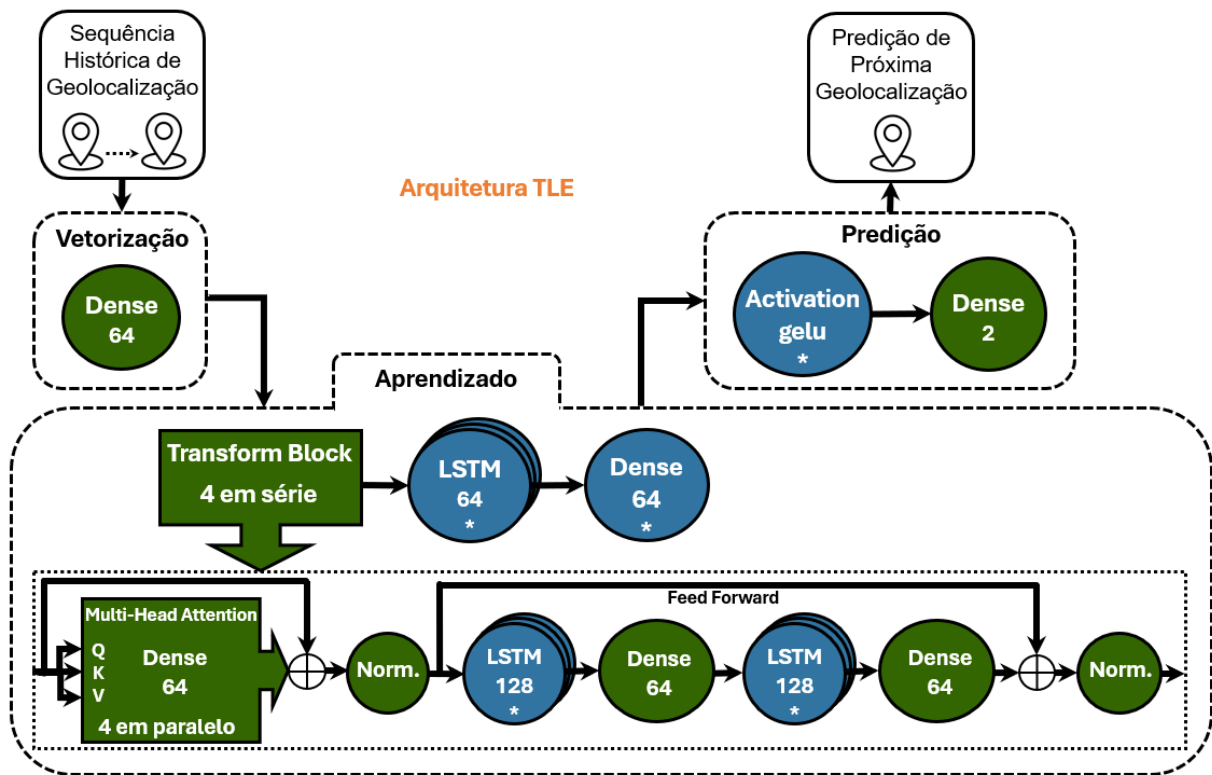


Figura 5.1: Funcionamento da arquitetura TLE implementada.

Block) é composto por uma estrutura de Atenção Multi-Cabeças (*Multi-Head Attention*) contendo quatro cabeças executando em paralelo, onde ocorre a função de “atenção” entre os pontos de cada sequência de entrada. Após a estrutura de Atenção Multi-Cabeças (*Multi-Head Attention*) ocorre a soma da saída dessa estrutura com sua entrada inicial (que, no primeiro bloco, é a saída da estrutura de Vetorização, e, nos blocos seguintes, é a saída do bloco anterior da série), operação conhecida como conexão residual (*residual connection*) seguido de uma camada de normalização dos dados. Na sequência da camada de normalização vem a estrutura de Propagação Direta (*Feed Forward*) também seguida de uma conexão residual e uma camada de normalização. Internamente da estrutura de Propagação Direta (*Feed Forward*) temos uma camada LSTM, seguida de uma camada Densa, mais uma camada LSTM, também seguida de uma camada Densa, culminando na conexão residual (*residual connection*) da estrutura de Propagação Direta (*Feed Forward*), seguida da camada de normalização e esta encerrando o processamento do Bloco Transformer (*Transform Block*) e entregando a sua saída para o próximo bloco ou camada. A estrutura de Aprendizado (*Learning*) então se completa seguindo com a camada LSTM e, na sequência, com mais uma camada Densa - que apesar de ser uma camada “genérica”, sua sequência após a camada LSTM foi inspirada na arquitetura LSTM em seu estado-da-arte empregada neste trabalho -, que entrega sua saída para a estrutura de Predição (*Predicting*).

Passando pelas estruturas de Vetorização (*Embedding*) e de Aprendizado (*Learning*), os dados processados nessas duas estruturas chegam na estrutura de Predição (*Predicting*). Essa estrutura de Predição (*Predicting*) abarca uma camada de Ativação (*Activation*), executando a função de ativação Unidade Linear de Erro Gaussiano (*Gaussian Error Linear Units - GELU*) (49) - que é mais uma peculiaridade da arquitetura TLE em relação à Transformer -, também inspirada na arquitetura LSTM em seu estado-da-arte, e, por

fim, uma camada Densa. A saída dessa última camada Densa é a saída final da arquitetura Transformer e LSTM Entrelaçados (TLE). Portanto, para realizar a predição, a camada de Ativação “GELU” realiza uma espécie de filtragem nos valores recebidos, aplicando uma transformação não linear suave que atenua valores negativos ou próximos a zero, para que a camada Densa final realize uma redução de dimensionalidade e extraia a predição das coordenadas de próxima geolocalização.

O termo Transformer e LSTM Entrelaçados (TLE) é cunhado pelo fato de haver a inserção de duas camadas LSTM intercaladas com as camadas densas originais expandindo a parte interna de Propagação Direta (*Feed Forward*) padrão da arquitetura Transformer original, conferindo uma capacidade de modelar dependências de curto e longo prazo também dentro da área de Propagação Direta (*Feed Forward*), adicionalmente à área de Atenção Multi-Cabeças (*Multi-Head Attention*). Além dessas duas camadas LSTM, após os Blocos Transformer (*Transform Blocks*), consolidando e sintetizando as informações dos Blocos Transformer (*Transform Blocks*) que contém a área de Atenção Multi-Cabeças (*Multi-Head Attention*) seguida da Propagação Direta (*Feed Forward*), há uma terceira camada LSTM, seguida de mais uma camada densa. Essa integração de camadas LSTM nas entranhas da arquitetura Transformer é o que confere a característica de entrelaçamento entre essas duas arquiteturas tão bem consolidadas na área de *deep learning*, criando a arquitetura Transformer e LSTM Entrelaçados (TLE) que combina mecanismos de atenção global com memória recorrente.

A arquitetura Transformer, por ser mais nova e mais complexa, foi escolhida como base principal da TLE e o entrelaçamento das camadas LSTM foi definido com base nas áreas da arquitetura Transformer onde faz-se sentido realizar sequências de camadas (*layers*) de redes neurais. Em seguida, tanto as posições exatas quanto os hiperparâmetros da arquitetura TLE foram definidos tomando por base suas arquiteturas de referência - Transformer e LSTM - e realizando testes com variações desses valores para alcançar o melhor desempenho. Assim apesar de ter sido testada apenas no contexto de predição de próximo sensor externo para veículos roubados, a arquitetura TLE, por ser uma arquitetura híbrida entrelaçada, tem potencial para se adaptar a diversos outros contextos e, quem sabe, incentivar o entrelaçamento de outras arquiteturas, dinamizando e acelerando a evolução de modelos de *deep learning*.

Outro aspecto a realçar é que em outros contextos de predição de sequências, como na área de Processamento de Linguagem Natural (PLN), sequências de entradas maiores costumam ser mais usadas e ter melhores resultados. Por outro lado, para predizer o próximo sensor externo de veículos roubados, quanto menor a necessidade de pontos para a sequência de entrada do modelo, mais rápido se conseguirá tentar abordar e recuperar o veículo. Desse modo, resta claro que essa nova arquitetura TLE, a qual tem apresentado vantagem nesse escopo específico - sequências de entradas curtas -, se mostra uma excelente opção ao contexto de predição de próximo sensor externo de veículos roubados. Nesse sentido, outras predições de sequências que venham a ter essa característica, de quanto menor a sequência de entrada melhor para o negócio, poderiam vir a se beneficiar também dessa inovação.

5.1 CONSIDERAÇÕES FINAIS

A arquitetura Transformer e LSTM Entrelaçados (TLE) e seu funcionamento foram apresentados em detalhando suas três principais estruturas (Vetorização (*Embedding*), Aprendizado (*Learning*) e Predição (*Predicting*)), com destaque para a inserção de 3 camadas *Long Short-Term Memory* (LSTM) em diferentes pontos da arquitetura Transformer, caracterizando, assim, o entrelaçamento entre essas duas arquiteturas de destaque em *deep learning*. Todas essas informações e características foram aplicadas e fazem parte de dois estudos que são detalhados nos próximos Capítulos 6 e 7.

6 PRIMEIRO ESTUDO - GANHO DE PREDIÇÃO

Neste capítulo serão retratados os detalhes do primeiro estudo, objetivando “Ganho de Predição”, o qual culminou na publicação do artigo (26) e recebeu o prêmio de Segundo Melhor Artigo do CSBC/SBCUP 2024 - XLIV Congresso da Sociedade Brasileira de Computação - XVI, durante o Simpósio Brasileiro de Computação Ubíqua e Pervasiva, da Sociedade Brasileira de Computação - SBC.

6.1 INTRODUÇÃO

Neste primeiro estudo, para uma adequada contextualização, contudo, evitando a repetitividade, vamos recapitular apenas um resumo, de forma consolidada, do cenário apresentado desse primeiro estudo no Capítulo 4 e na sequência detalhar mais especificidades deste primeiro estudo:

- **1º Estudo - Ganho de Predição.** Visão geral do problema: Problema de regressão. Insumos: Registros dos veículos ao passar por radares de velocidade. Bases de dados: Ceará (CE), Distrito Federal (DF) e Goiás (GO). Arquiteturas: LSTM, Transformer e TLE. Abordagem: “*inclusão da informação sobre qual via da rua o veículo foi identificado*”. Métricas: RMSE, MSE e MAE. Execuções: 54 cenários, 162 execuções de treinamentos e mais 162 execuções de predições.

Outrossim, mais especificamente, nesse primeiro estudo pretende-se responder às seguintes Questões de Pesquisa (QP) mencionadas no Capítulo 1: **QP1:** “*A informação sobre qual via da rua o veículo foi identificado auxilia na predição da próxima posição?*” e **QP2:** “*A nova arquitetura Transformer e LSTM Entrelaçados (TLE) trouxe ganho em relação às arquiteturas individuais de LSTM e Transformer?*”.

Nas próximas seções, serão definidos e apresentados o pré-processamento dos dados, as arquiteturas de IA, os experimentos com as métricas avaliadas, seguidos pelos resultados e análises e as considerações finais desse primeiro estudo.

6.2 PRÉ-PROCESSAMENTO

Para a realização desse primeiro estudo, em relação às bases de dados utilizadas neste estudo, cada arquitetura foi treinada com os registros de radares de velocidade que geraram alertas de veículos roubados ou furtados, durante o período de junho de 2021 a setembro de 2023, em três estados brasileiros: Ceará (CE), Goiás (GO) e Distrito Federal (DF).

Um aspecto fundamental para um treinamento e predição de sequências, especialmente no contexto de predição de próximo sensor externo de trajetória veicular é a definição da quantidade de pontos de geolocalização de entrada e de saída para o modelo de IA. Assim, para cada conjunto de arquitetura e estado brasileiro, o modelo foi treinado considerando três valores distintos de sequências para, com base nessas

sequências, o modelo prever a próxima geolocalização que o veículo roubado poderia ser identificado. Assim, foram realizados treinamentos com sequências históricas de 2 localizações para prever a 3ª, abordando um cenário minimalista; sequências de 10 localizações para prever a 11ª, abordando um cenário mediano e que - com base em testes iniciais exploratórios -, tende a ser o ponto ótimo de melhor desempenho de predição; e sequências de 30 localizações para prever a 31ª, abordando um cenário de maior volume histórico, para analisar a premissa de que quanto mais histórico, melhor é a predição.

Para aplicação da abordagem geral (*“inclusão da informação sobre qual via da rua o veículo foi identificado”*), duas abordagens específicas foram utilizadas para análise e comparação dos resultados sobre os registros dos sensores externos:

Abordagem 1 *Registros com identificação de qual via da rua (com ID do sensor externo) onde o veículo foi localizado, além da latitude e da longitude.*

Abordagem 2 *Registros sem identificação de qual via da rua (sem ID do sensor externo) onde o veículo foi localizado, utilizando apenas a latitude e a longitude.*

A função de perda (*loss function*) implementada no contexto deste estudo foi a “Root Mean Squared Error” (RMSE), que utiliza a distância euclidiana entre a predição e o ponto real. A RMSE foi adotada por obter resultados melhores, em razão de um melhor ajustes de pesos desse modelo de IA no contexto do *backpropagation*, em testes exploratórios em relação à “Mean Squared Error” (MSE), à “Mean Absolute Error” (MAE) e a outras funções customizadas que foram construídas e também testadas em uma fase exploratória inicial dos dados e modelos.

Em continuidade a este estudo, para padronizar as entradas dos três modelos e aumentar o volume de sequências reais disponíveis para os treinamentos, as sequências de registros de um mesmo veículo, quando menores que a sequência do treino, foram descartadas. Por outro lado, quando as sequências eram maiores que a sequência do treino, estas foram subdivididas utilizando o conceito de janela deslizante para se obter diversas sequências menores - com o tamanho exato da sequência de treino - a partir de uma sequência originalmente maior. O pseudocódigo que realiza essa subdivisão nas sequências da base de dados é apresentado no Algoritmo 1.

Com base em (50) onde se propõe uma divisão ótima de dados de treinamento e teste na proporção de $\sqrt{p} : 1$, onde p é o número de parâmetros do modelo, faz-se uma ponderação do resultado e divide-se em base de treinamento com 95% dos registros e base de teste com os 5% dos registros restantes. Essa divisão condiz com o contexto de alta variabilidade dos dados e modelos complexos de *deep learning*. A tabela 6.1 traz esses quantitativos.

Como boa prática em treinamentos de modelos de *deep learning*, os dados de entrada foram normalizados entre 0 e 1 (intervalo [0,1]) e os dados de teste também normalizados considerando apenas as estatísticas dos dados de treinamento. Em seguida, foi aplicado o conceito de janela deslizante e, então, as sequências de entrada foram embaralhadas randomicamente para melhorar a generalização do modelo.

Para analisarmos se a hipótese de que a informação de via da rua auxilia na predição, adota-se as Abordagens 1 e 2 para cada um desses cenários. Portanto, cada modelo foi treinado considerando os dados de ID do sensor externo e desconsiderando esse dado, para avaliar se essa informação melhora a

Algoritmo 1. Janela Deslizante.

```
1: Entrada: Conjunto de Registros R, número de registros em cada sequência n_seq
2: Inicialize listas vazias: input, output
3: Para  $i \in [n\_seq, \text{tamanho}(R))$  faça
4:   Se  $\forall j \in \text{subconjunto}(R[i-n\_seq+1, i])$ ,  $\text{identificador}(j) = \text{identificador}(R[i+1])$  então
5:      $input \leftarrow input + \text{subconjunto}(R[i-n\_seq, i])$ 
6:      $output \leftarrow output + R[i+1]$ 
7:   Fim se
8: Fim para
9: Retorne input, output
```

Fim do Algoritmo Janela Deslizante.

predição ou se, ao contrário, gera ruído no treinamento dificultando a predição. Esses IDs dos sensores são identificados por números inteiros, que guardam uma relação numérica de proximidade que, apesar de não ser totalmente proporcional, é aproveitada pela arquitetura em suas predição. Na figura 6.1 o diagrama ilustra a estrutura dos 18 treinamentos para cada arquitetura, totalizando 54 treinamentos, para gerar, então, as predições desses modelos treinados.

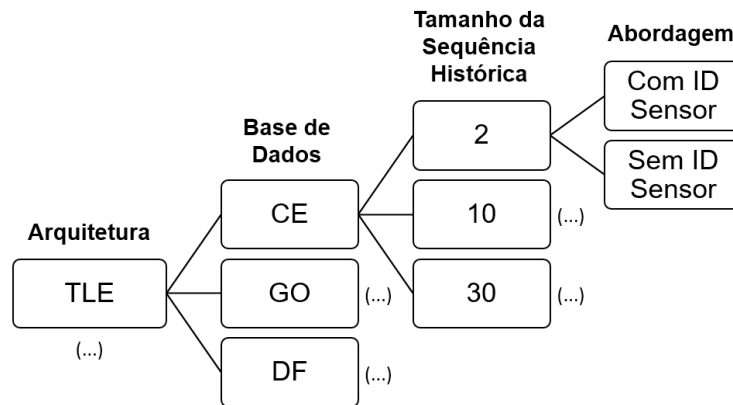


Figura 6.1: Diagrama dos treinamentos.

6.3 ARQUITETURAS COMPARADAS - 1º ESTUDO

As três arquiteturas utilizadas para os treinamentos e predições neste primeiro estudo foram: LSTM; Transformer; e Transformer e LSTM Entrelaçados (TLE). A seguir, as subseções 6.3.1, 6.3.2 e 6.3.3 detalham, respectivamente, as configurações e aspectos como quantidade de parâmetros treináveis para cada

Tabela 6.1: Quantidades de Predições.

Estado	Nº de Registros de Entrada	Nº de Predições de Treinamento	Nº de Predições de Teste	Total de Predições
CE	2	2120	106	2226
CE	10	920	46	966
CE	30	20	1	21
GO	2	7060	353	7413
GO	10	5200	260	5460
GO	30	3460	173	3633
DF	2	14240	712	14952
DF	10	5200	260	5460
DF	30	1920	96	2016

uma das três.

6.3.1 Arquitetura LSTM

A arquitetura *Long Short-Term Memory* (LSTM), como explicada no item 2.2.1.2 deste trabalho, é voltada para lidar com os desafios do aprendizado sequencial. A construção LSTM utilizada neste estudo segue o modelo em seu estado-da-arte descrito em (23) e (26), contando com características específicas como a função de ativação *Gaussian Error Linear Unit* (GELU) (49) e o uso do *Glorot Initialization* (51), para inicialização dos pesos na rede neural, e configurado com os seguintes parâmetros: otimizador AdamW (52), com taxa de aprendizado $learning_rate = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.98$ e $\epsilon = 10^{-8}$. Esse modelo de LSTM, quando criado, contém aproximadamente 3.4 milhões de parâmetros treináveis e está ilustrado na Figura 6.2.

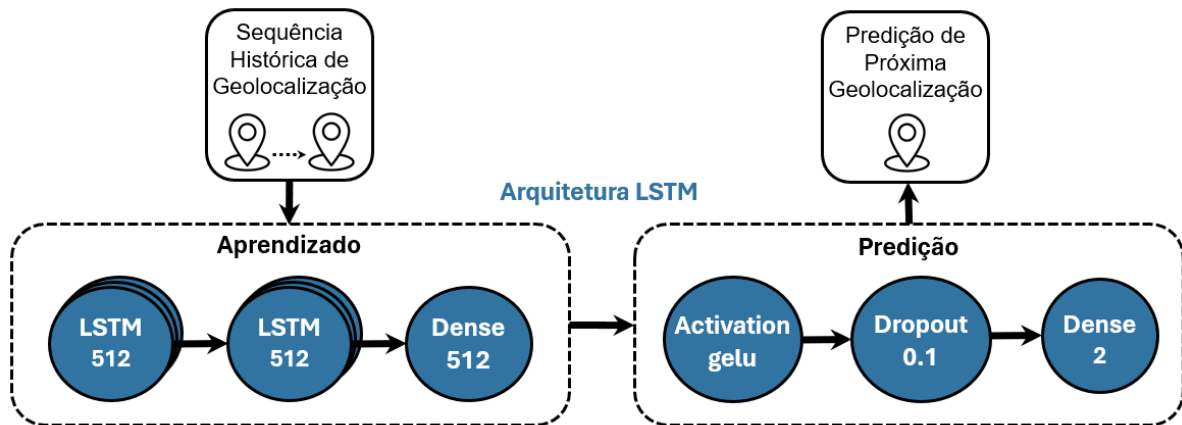


Figura 6.2: Arquitetura LSTM implementada.

6.3.2 Arquitetura Transformer

A arquitetura Transformer empregada neste estudo foi baseada em seu estado-da-arte descrito em (3), com as adaptações descritas no Capítulo 2, considerando que os tipos de dados de entrada não são palavras. Testes foram realizados e essa versão da arquitetura foi ajustada para sua melhor performance, contendo cerca de 100 mil parâmetros treináveis e está representada na Figura 6.3.

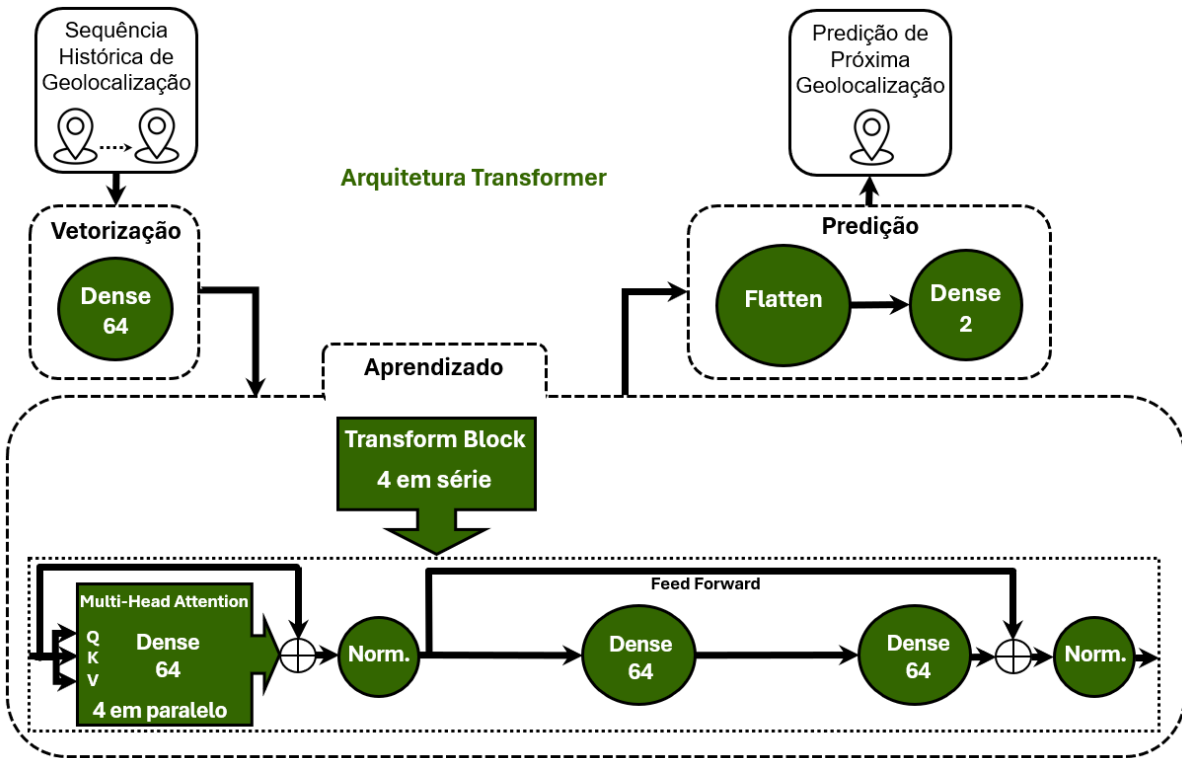


Figura 6.3: Arquitetura Transformer implementada.

Essa arquitetura Transformer por realizar execuções em paralelo internamente, diferentemente da LSTM que tem execução sequencial, juntando com a menor quantidade de parâmetros treináveis, termina por ter o tempo de treinamento significativamente menor em relação aos demais modelos deste trabalho.

6.3.3 Arquitetura TLE

A arquitetura Transformer e LSTM Entrelaçados (TLE) é o modelo inovador descrito no Capítulo 5. Por possuir características de atenção e recorrência, neste estudo, pretende-se avaliar o possível benefício de junção dessas técnicas. Neste estudo, a arquitetura TLE foi aplicada com a configuração padrão apresentada no Capítulo 5. Assim, quando criado, esse modelo possui aproximadamente 1 milhão de parâmetros treináveis e está ilustrado na Figura 5.1.

6.4 EXPERIMENTOS E ANÁLISES - 1º ESTUDO

Nesse primeiro estudo, foram implementadas três métricas de avaliação convencionais para problemas de regressão (44) (45) para analisar os resultados dos experimentos: RMSE (*Root Mean Squared Error*), MSE (*Mean Squared Error*) e MAE (*Mean Absolute Error*). Por ser um primeiro estudo e comparar uma nova arquitetura de predição, adotar três métricas que tendem a seguir a mesma tendência, ou seja, quando uma reduz as outras tendem a também reduzir e vice-versa, ao não encontrar grandes inconsistências, traz uma maior garantia de que não houve erro de execução nos experimentos. Assim, após todas as execuções previstas, a tabela 6.2 traz os dados de avaliação das predições para as três bases de teste (CE, GO e DF) em relação às métricas RMSE, MSE e MAE para cada arquitetura e seus respectivos 18 cenários.

Tabela 6.2: Resultados das avaliações das predições das bases de teste CE, GO e DF.

Parâmetros\Arquiteturas			LSTM			Transformer			TLE		
Estado	n_seq	Abord.	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE
CE	2	1	0.0072	5.36×10^{-5}	0.0052	0.0171	2.93×10^{-4}	0.0110	0.0112	1.26×10^{-4}	0.0084
	2	2	0.0074	5.59×10^{-5}	0.0053	0.0159	2.52×10^{-4}	0.0102	0.0120	1.45×10^{-4}	0.0086
	10	1	0.0070	4.86×10^{-5}	0.0053	0.0134	1.80×10^{-4}	0.0115	0.0096	9.15×10^{-5}	0.0080
	10	2	0.0076	5.72×10^{-5}	0.0057	0.0103	1.07×10^{-4}	0.0087	0.0102	1.06×10^{-4}	0.0089
	30	1	0.0102	1.03×10^{-4}	0.0102	0.0281	7.87×10^{-4}	0.0259	0.0154	2.37×10^{-4}	0.0153
	30	2	0.0120	1.45×10^{-4}	0.0107	0.0121	1.46×10^{-4}	0.0095	0.0138	1.91×10^{-4}	0.0119
GO	2	1	0.0076	1.20×10^{-4}	0.0026	0.0091	1.92×10^{-4}	0.0030	0.0058	6.85×10^{-5}	0.0024
	2	2	0.0075	1.19×10^{-4}	0.0026	0.0092	1.83×10^{-4}	0.0031	0.0064	8.68×10^{-5}	0.0024
	10	1	0.0024	5.90×10^{-6}	0.0017	0.0027	7.53×10^{-6}	0.0020	0.0024	5.97×10^{-6}	0.0017
	10	2	0.0029	8.74×10^{-6}	0.0023	0.0030	9.18×10^{-6}	0.0022	0.0028	8.05×10^{-6}	0.0020
	30	1	0.0025	6.52×10^{-6}	0.0018	0.0036	1.35×10^{-5}	0.0025	0.0036	1.31×10^{-5}	0.0025
	30	2	0.0026	6.67×10^{-6}	0.0018	0.0036	1.35×10^{-5}	0.0025	0.0037	1.37×10^{-5}	0.0025
DF	2	1	0.0899	0.0083	0.0654	0.0888	0.0081	0.0642	0.0890	0.0082	0.0640
	2	2	0.0900	0.0083	0.0654	0.0892	0.0082	0.0644	0.0894	0.0082	0.0645
	10	1	0.0798	0.0065	0.0582	0.0956	0.0095	0.0702	0.0780	0.0062	0.0568
	10	2	0.0801	0.0066	0.0584	0.1000	0.0104	0.0744	0.0867	0.0077	0.0619
	30	1	0.0694	0.0051	0.0463	0.0989	0.0098	0.0692	0.0855	0.0074	0.0633
	30	2	0.0697	0.0051	0.0463	0.1124	0.0127	0.0802	0.0859	0.0075	0.0570

Analisando os resultados e os gráficos das Figuras 6.4 e 6.5 percebe-se que a Abordagem 1 favorece uma melhor performance média para as 3 quantidades de sequências de entrada ($n_{seq} = (2, 10, 30)$) para predição de próximo sensor externo, havendo um ganho de 0,35% para $n_{seq} = 2$, calculando-se $1 - (28, 28/28, 38)$, um ganho de 10% para $n_{seq} = 10$ e um ganho de 9,1% para $n_{seq} = 30$. Para a média geral, a Abordagem 1 promove uma melhoria média de 6,5% no MSE das predições.

Em relação à análise entre as 3 arquiteturas, os gráficos das Figuras 6.6 e 6.7 mostram que a LSTM levou vantagem no MSE na maior parte das avaliações. Na média geral, a LSTM obteve uma performance

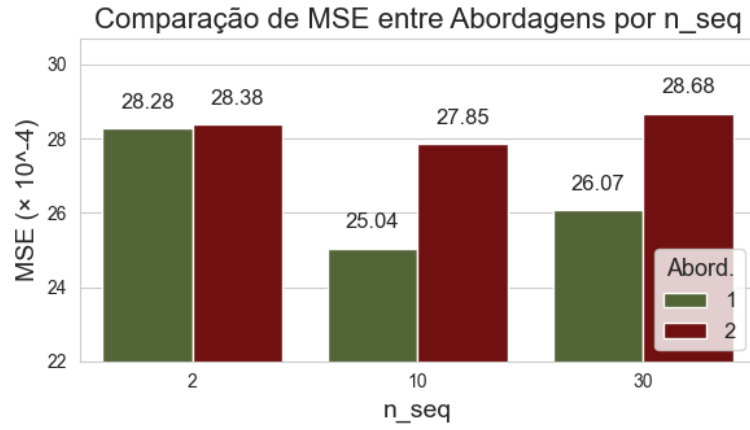


Figura 6.4: Comparação entre as abordagens 1 (com ID sensor) e 2 (sem ID sensor) em relação ao tamanho da sequência histórica de entrada.

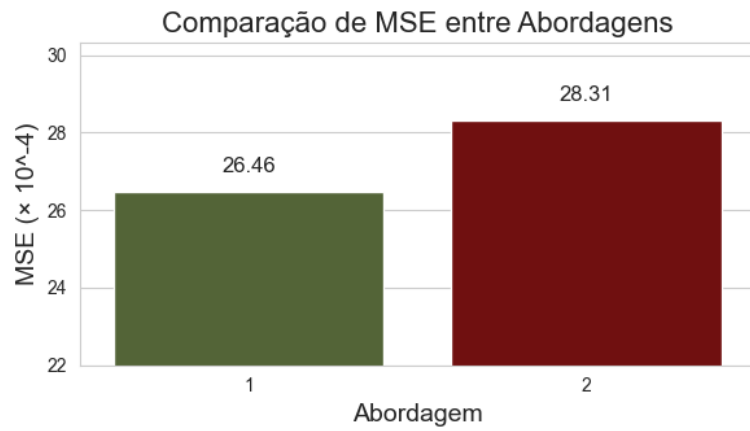


Figura 6.5: Comparação geral entre as abordagens 1 e 2.

12,4% melhor do que a TLE e 33,3% melhor do que a Transformer. No entanto, existe um cenário, que é para o $n_{seq} = 2$, em que a nova proposta de arquitetura TLE (Transformer e LSTM Entrelaçados) obtém o melhor MSE médio, sendo 0,74% melhor em comparação à LSTM e 2,3% melhor em comparação à Transformer. Considerando, ainda, que para $n_{seq} = 2$ observa-se uma tendência de minimização das variações de resultados e que mesmo na comparação entre as abordagens 1 e 2, que na média geral foi significativa (6,5%), porém para $n_{seq} = 2$ houve um ganho de apenas 0,35%, destaca-se que o ganho de 0,74% da TLE em relação à LSTM se torna ainda mais relevante, por ser mais que o dobro do ganho da abordagem 1 em relação à abordagem 2 para $n_{seq} = 2$.

Portanto, analisando a Questão de Pesquisa “**QP1**: A informação sobre qual via da rua o veículo foi identificado auxilia na predição da próxima posição?” e observando os resultados desse experimento, em especial as Figuras 6.4 e 6.5, percebe-se que em média há um benefício em utilizar a informação da via que o veículo foi identificado ao realizar os treinamentos e predições de modelos auxiliando na predição de próximo sensor externo de veículos roubados.

Em relação à Questão de Pesquisa “**QP2**: A nova arquitetura Transformer e LSTM Entrelaçados (TLE) trouxe ganho em relação às arquiteturas individuais de LSTM e Transformer?” observando os resultados

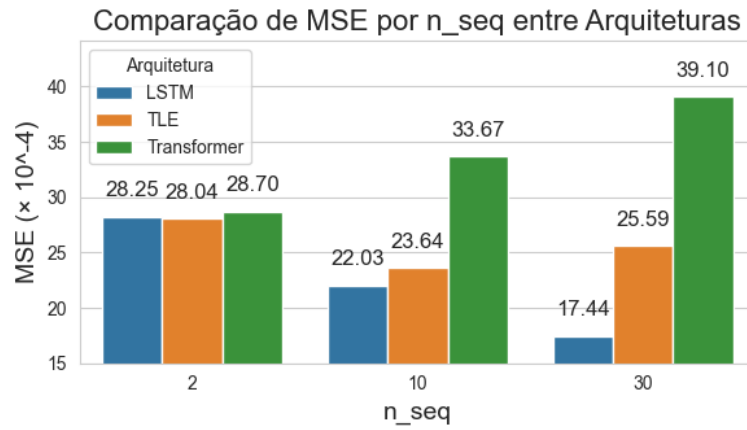


Figura 6.6: Comparação entre as arquiteturas LSTM, TLE e Transformer em relação ao tamanho da sequência histórica de entrada.

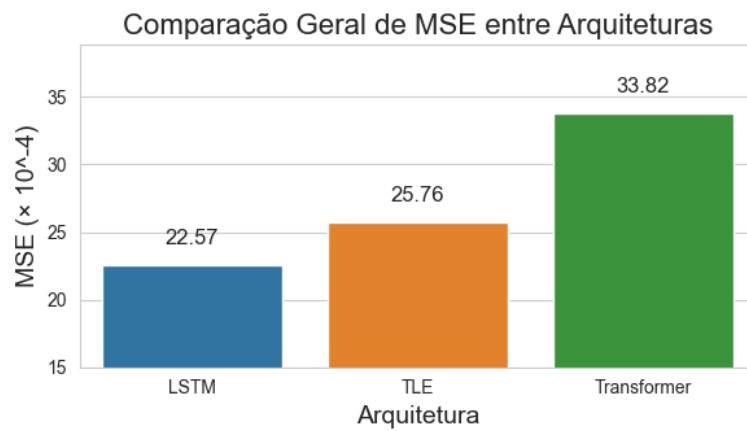


Figura 6.7: Comparação geral entre as arquiteturas LSTM, TLE e Transformer.

desse experimento, em especial as Figuras 6.6 e 6.7, percebe-se que não há uma vantagem absoluta da arquitetura TLE em relação à LSTM e à Transformer, no entanto, demonstra um potencial relevante de superação para predição de sequências históricas curtas, mais especificamente para 2 pontos históricos de entrada para a predição do terceiro ponto de geolocalização do veículo roubado, cabendo uma análise mais aprofundada sobre esse aspecto, procurando entender melhor qual o espectro de tamanho da sequência que a arquitetura TLE se destaca.

6.5 CONSIDERAÇÕES FINAIS

Os resultados mostraram que, em média, a informação de qual via o veículo foi identificado traz um efeito positivo na predição de próximo sensor externo, com um ganho médio geral do MSE de 6,5% em relação aos mesmos registros desconsiderando a informação de via da rua, identificando ganhos para os tamanhos de sequência 2 (0,35%), 10 (10%) e 30 (9,1%). Com isso entende-se que, mesmo em um contexto de comportamento anômalo do motorista, a informação da via da rua, aqui estudada, traz pistas de para

onde o veículo está indo, favorecendo a predição de próximo sensor externo.

Em relação à proposta da nova arquitetura TLE, não teve destaque amplo, mas atingiu resultado melhor na média geral do MSE para $n_{seq} = 2$, ou seja, 2 localizações de entrada para predizer a 3ª geolocalização. Mesmo com uma vantagem pequena (0,74% em relação à LSTM e 2,3% em relação à Transformer), mostra um potencial importante, valendo novos empregos e análises dessa arquitetura, pois, no contexto da predição de próximo sensor externo para veículos roubados, quanto menor a quantidade de localizações históricas para predizer a próxima, melhor para recuperar o veículo, minimizando as chances de uso do veículo roubado para cometer outros crimes subsequentes.

7 SEGUNDO ESTUDO - APOIO À DEFINIÇÃO DE MELHOR ARQUITETURA

Com a publicação do primeiro artigo apresentado no Capítulo 6, despertou-se interesse e curiosidade em relação aos achados, especialmente sobre a nova arquitetura e a necessidade de se investigar mais a fundo seu melhor escopo de atuação. Nesse contexto, a premiação do desse artigo trouxe o prestígio de sermos convidados para envio de uma versão estendida desse estudo a ser publicada no *Journal of Internet Services and Applications* (JISA). Essa situação nos encorajou a executarmos um segundo estudo, o qual é detalhado neste Capítulo.

7.1 INTRODUÇÃO

Uma parte do desafio de elaborar uma versão estendida de um estudo é produzir, ao menos, cerca de 40% de conteúdo novo, selecionando as ideias, e ainda manter a relação com o conteúdo anterior. Nesse sentido, como âncora ao primeiro artigo, mantivemos o foco em predição de próximo sensor externo para veículos roubados e incorporamos a conceito de incluir a informação sobre qual via da rua o veículo foi identificado em todos os experimentos desse segundo estudo. Por outro lado, como novidade, trouxemos a arquitetura *Extreme Gradient Boosting* (XGBoost) para a comparação, utilizamos uma nova base de dados do estado de São Paulo (SP), aplicamos um novo espectro de variação do tamanho da sequência histórica de pontos (1 a 8) e novas abordagens foram analisadas. Adicionalmente, para mensuração dos resultados, mantivemos a métrica Erro Médio Quadrático (*Mean Squared Error* - MSE) e adicionamos a métrica Coeficiente de Determinação (R^2). Assim, vamos recapitular apenas um resumo, de forma consolidada, do cenário apresentado desse segundo estudo no Capítulo 4 e na sequência detalhar suas especificidades:

- **2º Estudo - Apoio à Definição de Melhor Arquitetura.** Visão geral do problema: Programa de regressão. Insumos: Registros dos veículos ao passar por radares de velocidade. Base de dados: São Paulo (SP). Arquiteturas: XGBoost, LSTM e TLE. Abordagens: “*identificação do ponto de transição entre as melhores arquiteturas de predição de próximo sensor externo*” e “*identificação do impacto da variação do formato de representação geográfica das geolocalizações nas predições e no ponto de transição entre as melhores arquiteturas de predição de próximo sensor externo*”. Métricas: MSE e R^2 . Execuções: 42 cenários, 84 execuções de treinamentos e mais 84 execuções de predições.

Desse modo, além de considerar a influência das mudanças na representação geográfica, este segundo estudo pretende analisar arquiteturas para predição de sequências com poucos pontos históricos, se concentrando nessas três principais arquiteturas em seu estado-da-arte: LSTM, TLE e XGBoost. Outrossim, em continuidade ao primeiro estudo, neste segundo estudo, objetivando o “Apoio à Definição de Melhor Arquitetura”, buscamos, mais especificamente, responder às seguintes Questões de Pesquisa (QP): **QP3:**

Qual arquitetura apresenta melhor desempenho com comprimentos de sequência curtos para predição?; QP4: Existe um ponto de transição onde se altera qual a arquitetura com melhor desempenho?; QP5: O formato da representação geográfica impacta os resultados da predição?; QP6: O formato da representação geográfica pode alterar o ponto de transição entre as arquiteturas com melhor desempenho, caso exista?; e QP7: A nova arquitetura Transformer e LSTM Entrelaçados (TLE) é vantajosa?.

A QP7 trata-se de uma continuidade da QP2 “A nova arquitetura Transformer e LSTM Entrelaçados (TLE) trouxe ganho em relação às arquiteturas individuais de LSTM e Transformer?” para uma confirmação e análise mais aprofundada dos achados do primeiro estudo. Ademais, além das informações sobre a faixa de rodagem do veículo, outro fator que pode impactar os resultados da predição é a forma como a geolocalização do sensor externo é representada. Por exemplo, para o mesmo ponto geográfico, em vez de usar latitude e longitude, é possível empregar *Northing* e *Easting*, potencialmente utilizando diferentes sistemas de referência. Isso ocorre porque os modelos de IA, especialmente em problemas de regressão como este, processam dados numéricos e, dependendo das relações inerentes a cada representação geográfica, bem como da distribuição e das distâncias entre os registros de geolocalização, o modelo pode capturar padrões nas sequências de dados de entrada de forma mais eficaz, igualmente ou de maneira menos eficiente.

Nas próximas seções, serão definidos e apresentados o pré-processamento dos dados, as arquiteturas de IA, os experimentos com as métricas empregadas, seguidos pelos resultados e análises e as considerações finais desse segundo estudo.

7.2 PRÉ-PROCESSAMENTO

Para este segundo estudo, utilizou-se um banco de dados contendo registros de radares de velocidade associados a roubos e furtos de veículos no estado de São Paulo (SP), Brasil, abrangendo o período de 30 de abril de 2021 a 23 de maio de 2024. Este conjunto de dados inclui placas de veículos anonimizadas, carimbos de tempo (*timestamps*) e códigos únicos de identificação dos radares de velocidade (ID do sensor externo). Ao integrar essas informações a um segundo banco de dados contendo as coordenadas de latitude e longitude correspondentes a cada código único de radar de velocidade, foi possível determinar a geolocalização de cada registro.

Um dos objetivos deste estudo é investigar o impacto das mudanças na representação geográfica. Portanto, além das informações de latitude e longitude, os dados foram transformados, criando novas colunas na base de dados com dois formatos adicionais de representação geográfica para os mesmos pontos correspondentes de latitude e longitude. As três representações geográficas, com diferenças de precisão e projeção, são descritas no item 2.2.2 deste trabalho e, adicionando essas transformações, foram utilizadas as seguintes representações:

- **EPSG:4326**, também conhecido como Latitude e Longitude na versão mais recente do World Geodetic System 1984 (WGS84), que utiliza coordenadas geográficas latitude e longitude em graus;
- **EPSG:3857**, também conhecido como Web Mercator, que emprega as medidas *Easting* e *Northing* em metros, em vez de graus como na latitude e longitude;

- **EPSG:32723**, também conhecido como Universal Transverse Mercator (UTM), que também utiliza *Easting* e *Northing* em metros, mas com sistemas de referência distintos.

Para abordar a **QP3** (*Qual arquitetura apresenta o melhor desempenho para predições com sequências curtas?*) e a **QP4** (*Existe um ponto de transição onde a arquitetura de melhor desempenho muda?*), e com base nos achados de (26), onde TLE e LSTM apresentaram uma mudança de desempenho entre comprimentos de sequência de 2 a 10, este estudo, utilizando a representação geográfica WGS84, iniciou os experimentos com comprimento de sequência igual a 1 e aumentou progressivamente para encontrar um possível ponto de transição e garantir que a mudança de desempenho não fosse revertida. Esse processo abrangeu comprimentos de sequência de 1 a 8 pontos históricos, treinando, prevendo e adicionando um ponto por vez.

Em seguida, para responder à **QP5** (*O formato da representação geográfica impacta os resultados da predição?*) e à **QP6** (*O formato da representação geográfica pode alterar o ponto de transição entre as arquiteturas de melhor desempenho, caso exista?*), o estudo treinou e previu utilizando os comprimentos de sequência nos limites do ponto de transição identificado (observado em **QP3** e **QP4**), mas desta vez utilizando as representações Web Mercator e UTM. O objetivo principal foi determinar se a representação geográfica influenciava consistentemente os resultados, enquanto o objetivo secundário foi avaliar se o ponto de transição permanecia entre os mesmos comprimentos de sequência, identificados como 5 e 6.

Para avaliar os resultados, foram utilizadas as métricas de Erro Quadrático Médio (*Mean Squared Error* - MSE) e Coeficiente de Determinação (*R-Squared* - R^2). O MSE é uma métrica padrão para problemas de regressão (44) (45), enquanto o R^2 , ou coeficiente de determinação, é amplamente utilizado no aprendizado de máquina como métrica de avaliação para modelos de regressão, pois ajuda a determinar o grau de ajuste do modelo aos dados. O valor de R^2 varia de 0 a 1, onde 0 indica que o modelo não explica nenhuma variação, e 1 indica que o modelo explica toda a variação observada (46) (47). Assim, para o MSE, valores menores indicam melhor desempenho, e para o R^2 , valores maiores indicam melhor desempenho.

Os dados foram divididos em 80% para treinamento e 20% para teste, uma divisão comum no treinamento de modelos de aprendizado de máquina (53) (54). Diferentemente de estudos como (19) e (26), que utilizaram janelas deslizantes para aumentar o número de sequências, este estudo, com maior foco na **QP3** (*Qual arquitetura apresenta o melhor desempenho para predições com sequências curtas?*), utilizou apenas os primeiros n_{seq} registros de cada veículo para cada dia como sequência de entrada para os modelos, onde n_{seq} representa o comprimento da sequência. Além disso, para garantir uma análise de resultados mais robusta, os dados foram primeiramente ordenados pelo identificador único do veículo e, em seguida, pelo timestamp, reduzindo possíveis vieses temporais. As avaliações foram realizadas duas vezes para cada conjunto de (arquitetura, n_{seq} , representação geográfica): primeiro utilizando os últimos 20% dos dados como conjunto de teste e, depois, utilizando os primeiros 20%, sendo o valor final de cada métrica obtido como a média dos resultados em cada teste. Essa abordagem é semelhante a uma validação cruzada parcial do tipo *k-fold*, garantindo uma análise robusta e eficiente.

7.3 ARQUITETURAS COMPARADAS - 2º ESTUDO

Três arquiteturas foram empregadas para treinamento e predição deste segundo estudo: *Extreme Gradient Boosting* (XGBoost), *Long Short-Term Memory* (LSTM) e Transformer e LSTM Entrelaçados (TLE). Essas arquiteturas são apresentadas e detalhadas nas subseções 7.3.1, 7.3.2 e 7.3.3, respectivamente.

7.3.1 Arquitetura XGBoost

A arquitetura de Aumento Extremo de Gradiente (*Extreme Gradient Boosting* - XGBoost) utilizada foi baseada na arquitetura GBRT descrita no item 2.2.1.1 deste trabalho e configurada com os seguintes parâmetros, detalhados no artigo (13): $max_depth = 3$; $learning_rate = 1$; $n_estimators = 100$; $min_child_weight = 1$; e $scale_pos_weight = 1$.

Durante o entendimento da arquitetura e sua implementação, percebemos que para a arquitetura XGBoost, a entrada do modelo precisa ser em formato tabular (matriz de atributos), onde cada linha representa um registro e cada coluna representa um atributo (*feature*), não havendo suporte “nativo” para sequências 2D como ocorre nas arquiteturas LSTM e Transformers. Dessa forma, como este estudo trabalha com sequências de duas dimensões, como, por exemplo, as coordenadas de latitude e longitude de cada ponto da sequência histórica, uma instância desse modelo foi usada para prever a parte 1, como latitude ou *Northing*, e outra instância usada para prever a parte 2, como longitude ou *Easting*. Assim, esta arquitetura XGBoost, quando criadas as duas instâncias, consiste em aproximadamente 1,4 mil divisões (*splits*), o equivalente a cerca de 1,4 mil parâmetros treináveis, e está ilustrada na Figura 7.1.

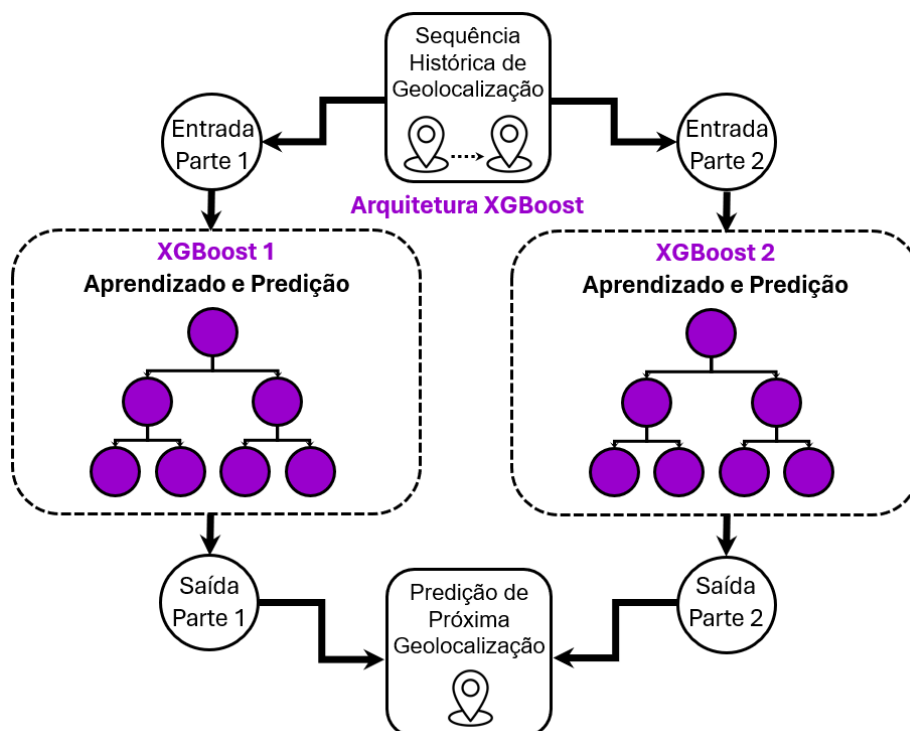


Figura 7.1: Arquitetura XGBoost implementada.

7.3.2 Arquitetura LSTM

A arquitetura *Long Short-Term Memory* (LSTM) utilizada neste segundo estudo segue o modelo em seu estado-da-arte descrito em (23) e (26) e apresentado no item 6.3.1, auxiliando na comparação de resultados com o primeiro estudo e contando, também, com aproximadamente 3.4 milhões de parâmetros treináveis, sendo ilustrada na Figura 6.2.

7.3.3 Arquitetura TLE

A arquitetura Transformer e LSTM Entrelaçados (TLE) aplicada neste segundo estudo é a mesma do primeiro estudo e está descrita no Capítulo 5 deste trabalho, também atingindo cerca de 1 milhão de parâmetros treináveis e ilustrada na Figura 5.1. No entanto, a arquitetura será analisada em novos cenários, dados e métrica, tornando os resultados gerais sobre a arquitetura TLE ainda mais robustos.

7.4 EXPERIMENTOS E ANÁLISES - 2º ESTUDO

Dado que o processo de aprendizado de máquina é altamente experimental (55), a função de perda *Root Mean Squared Error* (RMSE) foi adotada para as arquiteturas LSTM e TLE, pois apresentou melhores resultados em diversos testes de análise exploratória em comparação com *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE) e outras métricas personalizadas testadas. No entanto, para o XGBoost, a função de perda utilizada foi o *Mean Squared Error* (MSE), que é seu padrão e melhor opção para esse contexto, obtendo melhores resultados.

Para a análise dos resultados, este estudo investiga quatro principais questões de pesquisa, com base nos resultados obtidos a partir de três arquiteturas (LSTM, XGBoost e TLE) combinadas com três representações geográficas (WGS84, Web Mercator e UTM). O objetivo foi avaliar o desempenho de cada arquitetura em diferentes comprimentos de sequência curta e formatos de representação geográfica, bem como identificar possíveis pontos de transição entre as arquiteturas de melhor desempenho.

Para a primeira questão de pesquisa, **QP3: Qual arquitetura apresenta melhor desempenho com sequências curtas para predição?**, os modelos foram testados com comprimentos de sequência variando de 1 a 8, utilizando a representação geográfica WGS84. Os resultados indicaram que a arquitetura TLE apresentou o melhor desempenho para os menores comprimentos de sequência (1 a 5), enquanto a LSTM começou a superar a TLE à medida que o comprimento da sequência aumentou além de 5 (6 a 8). Isso sugere que a TLE é mais eficaz na captura de informações de sequências curtas, enquanto a LSTM se beneficia de dados históricos mais longos. A arquitetura XGBoost apresentou desempenho consistentemente inferior em comparação com a LSTM e a TLE em todos os cenários analisados. A Tabela 7.1 apresenta esses resultados utilizando as métricas MSE e R^2 .

Adicionalmente, para a segunda questão de pesquisa, **QP4: Existe um ponto de transição no qual a arquitetura de melhor desempenho muda?**, em conformidade com os achados de (26), a TLE apresentou melhor desempenho para sequências de comprimento 1 a 5, mas um ponto de transição foi identificado entre os comprimentos 5 e 6, onde a LSTM começou a superar a TLE. O XGBoost apresentou desempenho

Tabela 7.1: Comparação dos valores de MSE e R^2 para diferentes comprimentos de sequência nas arquiteturas XGBoost, LSTM e TLE.

Comprimento da Sequência	XGBoost		LSTM		TLE	
	MSE	R^2	MSE	R^2	MSE	R^2
1	0,0004848	0,9034	0,0004809	0,9047	0,0004741	0,9059
2	0,0003932	0,9104	0,0003841	0,9141	0,0003473	0,9217
3	0,0003845	0,9061	0,0003468	0,9172	0,0003353	0,9197
4	0,0003725	0,9083	0,0003490	0,9150	0,0003237	0,9208
5	0,0004016	0,8995	0,0003278	0,9153	0,0003121	0,9200
6	0,0004310	0,8918	0,0002852	0,9223	0,0002948	0,9193
7	0,0004227	0,8797	0,0002884	0,9186	0,0003099	0,9127
8	0,0004395	0,8788	0,0003001	0,9114	0,0003286	0,9030

consistentemente inferior em relação às outras duas arquiteturas em ambas as faixas de comprimento de sequência, de 1 a 5 e de 6 a 8. As Figuras 7.2 e 7.3 também demonstra esse ponto de transição nos resultados, utilizando a representação geográfica WGS84 junto às métricas MSE e R^2 , respectivamente, mostrando onde a curva da LSTM cruza a curva da TLE entre os comprimentos de sequência 5 e 6.

Para responder à **QP5**: *O formato de representação geográfica impacta os resultados da predição?*, além da representação geográfica WGS84, a performance de cada arquitetura foi avaliada convertendo as geolocalizações de WGS84 para as representações Web Mercator e UTM. Utilizando as mesmas métricas, MSE e R^2 , este estudo compara Web Mercator em relação ao WGS84, UTM em relação ao WGS84 e UTM em relação ao Web Mercator. Para interpretar os gráficos, é essencial observar que valores mais baixos de MSE indicam melhor desempenho, assim como valores mais altos de R^2 . Considerando esses fatores e analisando as comparações para cada arquitetura, identificamos três possíveis cenários para cada métrica ao comparar os resultados de duas representações geográficas distintas:

1. MSE da representação A dividido pelo MSE da representação B $< 1,0$: Indica que a representação A apresenta vantagem sobre a representação B em termos de MSE;
2. MSE da representação A dividido pelo MSE da representação B $> 1,0$: Indica que a representação A apresenta desvantagem em relação à representação B em termos de MSE;
3. MSE da representação A dividido pelo MSE da representação B $= 1,0$: Indica que a mudança na representação geográfica não teve impacto no MSE;
4. R^2 da representação A dividido pelo R^2 da representação B $> 1,0$: Indica que a representação A apresenta vantagem sobre a representação B em termos de R^2 ;
5. R^2 da representação A dividido pelo R^2 da representação B $< 1,0$: Indica que a representação A apresenta desvantagem em relação à representação B em termos de R^2 ;

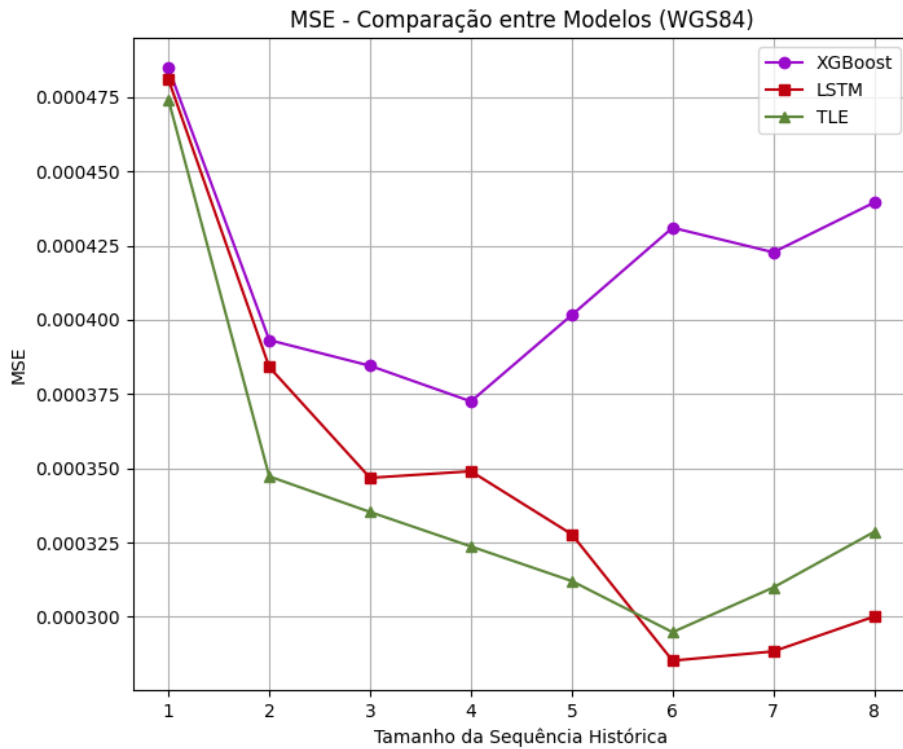


Figura 7.2: Comparação das arquiteturas XGBoost, LSTM e TLE utilizando a métrica MSE.

6. R^2 da representação A dividido pelo R^2 da representação B = 1,0: Indica que a mudança na representação geográfica não teve impacto no R^2 .

A Tabela 7.2 mostra que, em todas as comparações, nenhuma das conversões de representação geográfica manteve os mesmos valores de MSE ou R^2 . Esse achado indica que o formato de representação da geolocalização impacta os resultados. Embora essa não fosse uma questão de pesquisa primária, analisamos adicionalmente se o impacto foi positivo ou negativo, uma vez que permanece um aspecto intrigante e relevante do estudo.

Para essa avaliação detalhada, é essencial combinar as métricas MSE e R^2 para garantir análises e conclusões consistentes. Se uma mudança na representação geográfica levou a melhorias tanto no MSE quanto no R^2 , a mudança foi classificada como positiva. Por outro lado, se a mudança resultou em quedas em ambas as métricas, ela foi considerada negativa. Finalmente, se uma métrica melhorou enquanto a outra piorou, independentemente de qual, o impacto foi classificado como inconclusivo. Utilizando essa abordagem e analisando Figura 7.4, Figura 7.5 e Figura 7.6, que ilustram o impacto da mudança de cada representação geográfica, podemos construir a síntese apresentada na Tabela 7.3, que resume os resultados das comparações.

A análise indicou que a mudança para o formato Web Mercator não demonstrou um impacto consistentemente positivo ou negativo neste estudo, enquanto a adoção do UTM em relação ao WGS84 ou ao Web Mercator teve um efeito positivo no XGBoost e no LSTM para os três comprimentos de sequência analisados (4, 5 e 6). Para o TLE, o UTM teve um impacto positivo para $n_{seq} = 4$, mas um impacto negativo para os comprimentos de sequência 5 e 6. Esses resultados demonstram que a representação geo-

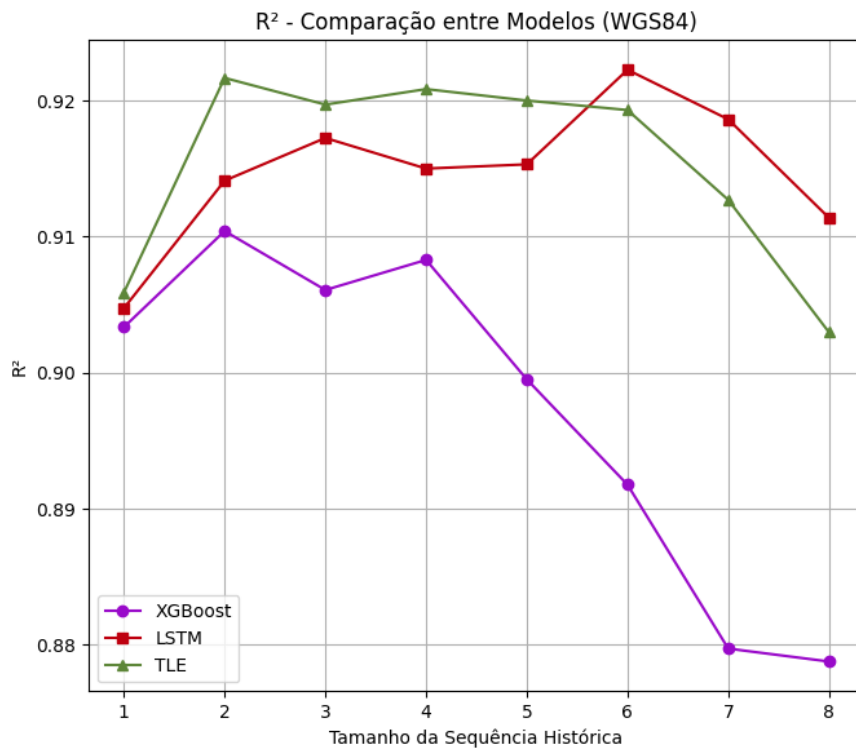


Figura 7.3: Comparação das arquiteturas XGBoost, LSTM e TLE utilizando a métrica R^2 .

gráfica impacta os resultados da predição e que o efeito pode variar conforme o comprimento da sequência, resultando em uma influência positiva ou negativa.

Para responder à sexta questão de pesquisa, **QP6**: *O formato de representação geográfica pode alterar o ponto de transição entre as arquiteturas com melhor desempenho, caso exista?*, este estudo analisou o ponto de transição identificado nas questões **QP3** e **QP4**. Utilizando a mesma faixa de comprimentos de sequência (4, 5 e 6), foi possível observar que, com o formato de representação WGS84, o ponto de transição entre LSTM e TLE, que inicialmente ocorria entre os comprimentos de sequência 5 e 6, foi deslocado para entre os comprimentos 4 e 5 com a adoção do formato de representação UTM. Com base nesses resultados, entende-se que, dependendo da representação geográfica utilizada, o ponto de transição entre as arquiteturas com melhor desempenho pode, de fato, variar, conforme ilustrado na Figura 7.7.

Esses resultados mostram impactos intrigantes ao variar a representação geográfica, pois o ponto de transição é afetado não apenas porque uma arquitetura se beneficiou mais do que outra ao mudarmos o formato da representação geográfica. No estudo, ao analisarmos o caso específico de comparação entre os formatos UTM e WGS84 na mudança do ponto de transição, a arquitetura TLE tem ganho para $n_{seq} = 4$, onde essa arquitetura permanece superior à LSTM, no entanto, para $n_{seq} = 5$ e $n_{seq} = 6$ o resultado da mudança no formato da representação gera uma perda de desempenho para a TLE, enquanto continua a resultar em ganho de desempenho para as arquiteturas LSTM e XGBoost. Isso colabora para o deslocamento do ponto de transição inicialmente entre os tamanhos de sequência 5 e 6 para entre os tamanhos de sequência 4 e 5. Já para a variação com o formato do Web Mercator, não houve alteração do ponto de transição, provavelmente pelo maior grau de impactos inconsistentes em sua aplicação, por essa razão não foi gerado o gráfico comparativo dessa mudança de representação relacionada ao ponto de transição.

Tabela 7.2: Comparação das mudanças nos formatos de representação geográfica.

Representação Comparação	Tamanho da Sequência	XGBOOST		LSTM		TLE	
		MSE	R ²	MSE	R ²	MSE	R ²
Web Mercator / WGS84	4	0,9940054	1,0007	1,0036207	1,0001	1,0160543	0,9989
Web Mercator / WGS84	5	1,0040891	1,0002	1,0116758	0,9993	1,0040741	1,0009
Web Mercator / WGS84	6	1,0950706	0,993	1,0023664	1,0003	0,9985159	1,0008
UTM / WGS84	4	0,9630129	1,0015	0,9769478	1,0005	0,9723658	1,0009
UTM / WGS84	5	0,9333241	1,0037	0,9772796	1,0009	1,0482563	0,9932
UTM / WGS84	6	0,9882064	1,0001	0,9724874	1,0011	1,0024658	0,9995
UTM / Web Mercator	4	0,9688205	1,0008	0,9734234	1,0005	0,9570018	1,0021
UTM / Web Mercator	5	0,9295231	1,0035	0,9660008	1,0015	1,044003	0,9924
UTM / Web Mercator	6	0,9024135	1,0072	0,9701915	1,0007	1,0039557	0,9987

Tabela 7.3: Impacto das mudanças relativas nos formatos de representação geográfica ao longo dos comprimentos de sequência.

Representação Geográfica Comparação	Impacto no XGBoost			Impacto no LSTM			Impacto no TLE		
	n_seq=4	n_seq=5	n_seq=6	n_seq=4	n_seq=5	n_seq=6	n_seq=4	n_seq=5	n_seq=6
Web Mercator / WGS84	Positivo	Inconclusivo	Negativo	Inconclusivo	Negativo	Inconclusivo	Negativo	Inconclusivo	Positivo
UTM / WGS84	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo	Negativo	Negativo
UTM / Web Mercator	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo	Negativo	Negativo

Para responder à última questão de pesquisa, **QP7: A nova arquitetura Transformer e LSTM Entrelaçados (TLE) é vantajosa?** é importante compararmos os resultados deste estudo com pesquisas anteriores. Neste estudo observa-se que a arquitetura Transformer e LSTM Entrelaçados (TLE) apresenta melhor desempenho para sequências curtas - abaixo de 6 utilizando WGS84 e abaixo de 5 adotando UTM -, enquanto a LSTM se torna mais eficaz para comprimentos de sequência maiores. Comparando com os resultados anteriores, esse achado está alinhado com (26), que indicou haver um ponto de transição entre essas arquiteturas para determinados comprimentos de sequência, haja vista a TLE ter obtido melhor desempenho para o comprimento de sequência histórica de dois pontos, enquanto a LSTM ter superado a TLE para dez pontos de sequência histórica. Considerando, ainda, que no contexto de veículos roubados é vantajoso utilizar menos pontos de sequência histórica, entende-se que a arquitetura Transformer e LSTM Entrelaçados (TLE) é vantajosa para predição utilizando sequências curtas.

Em termos de representação geográfica, este estudo fornece uma análise mais detalhada e abrangente sobre como diferentes sistemas de coordenadas impactam a precisão da predição, um tema menos explorado em pesquisas anteriores, como (13). Essas comparações são apresentadas na Tabela 7.1 e na Tabela 7.2, e ilustradas nas Figuras 7.4, 7.5 e 7.6, sintetizando o desempenho das arquiteturas XGBoost, LSTM e TLE em diferentes representações geográficas e comprimentos de sequência.

Diferentemente de trabalhos anteriores que focam exclusivamente no uso de diferentes arquiteturas ou na inclusão de informações adicionais, este estudo contribui significativamente ao demonstrar que tanto o comprimento da sequência quanto o formato de representação geográfica desempenham papéis essenciais na definição da arquitetura ideal para predição de trajetórias veiculares, em especial para veículos roubados.

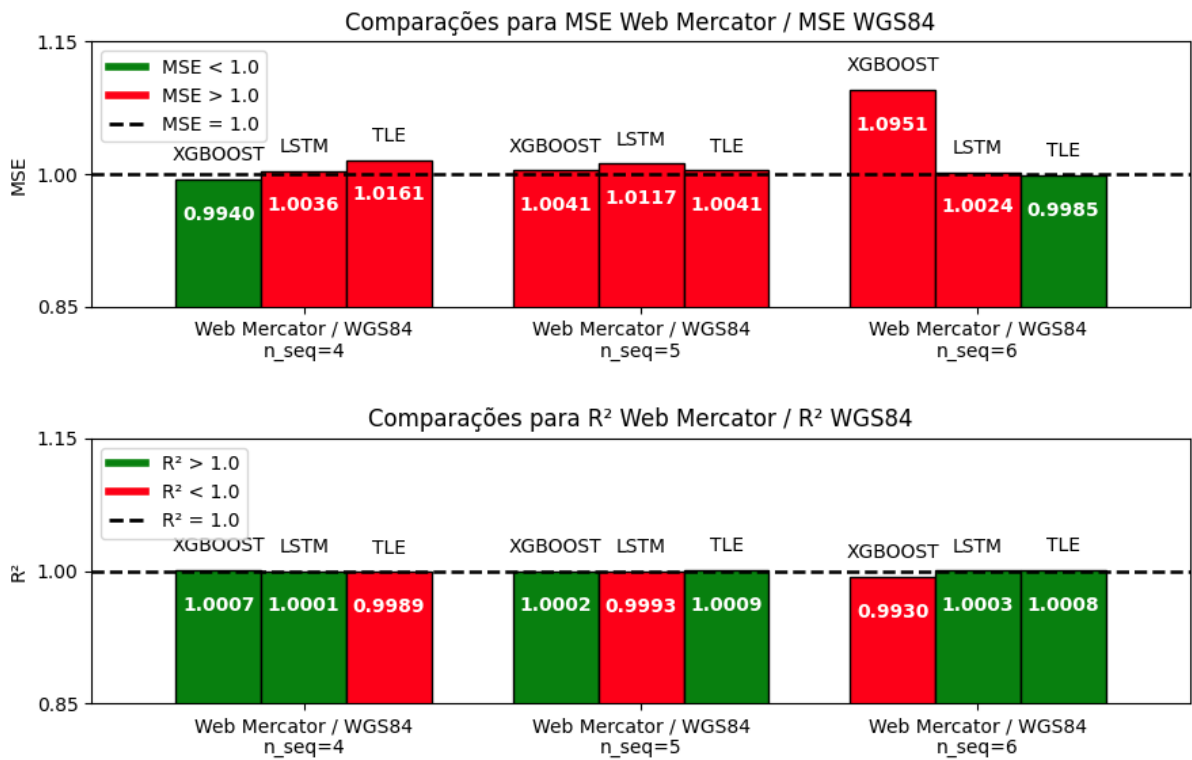


Figura 7.4: Comparação do formato de representação geográfica Web Mercator sobre WGS84 utilizando as métricas MSE e R².

7.5 CONSIDERAÇÕES FINAIS

Este segundo estudo trouxe uma análise mais aprofundada, como se estivéssemos olhando com uma lupa, a questão de ter a arquitetura TLE como mais vantajosa para determinado escopo e a arquitetura LSTM para outro. Nesse sentido, foram mantidas essas duas arquiteturas do primeiro estudo e adicionada mais uma diferente para agregar a comparação. De maneira similar, foi mantida a métrica MSE e adicionada a métrica R². Não obstante, foi mantida um tamanho de sequência do primeiro estudo ($n_{seq} = 2$) e adicionados novos tamanhos (1 a 8). Para completar de forma surpreendente essa análise, foi mantida a representação geográfica de latitude e longitude e avaliadas mais duas diferentes representações geográficas (Web Mercator e UTM). Dessa forma, este segundo estudo conseguiu expandir de forma significativa as perspectivas em relação à essa nova arquitetura TLE e suas concorrentes diretas.

Os dois estudos trouxeram contribuições além da análise específica da arquitetura TLE. Enquanto no primeiro estudo houve ganho de desempenho na predição de veículos roubados, neste segundo estudo, objetivando o “Apoio à Definição de Melhor Arquitetura”, demonstrou-se a relevância em avaliar dois critérios fundamentais para a escolha da arquitetura de melhor desempenho na predição do próximo sensor externo no contexto de veículos roubados: o comprimento da sequência histórica e o formato de representação da geolocalização.

Os resultados destacam a importância desses dois critérios, dado que foi identificado um ponto de transição no qual a arquitetura ótima muda dependendo do comprimento da sequência histórica. Além disso,

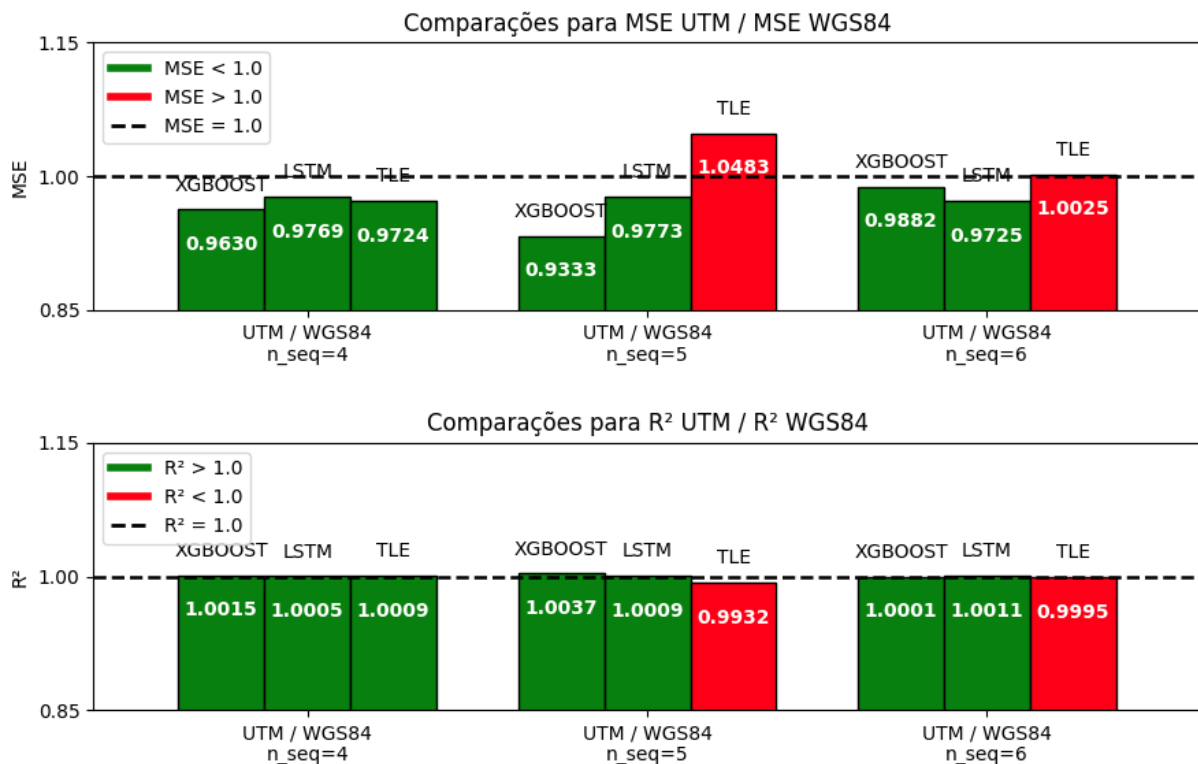


Figura 7.5: Comparação do formato de representação geográfica UTM sobre WGS84 utilizando as métricas MSE e R^2 .

observou-se que alterações no formato de representação geográfica podem ter impacto positivo, negativo ou inconclusivo nas predições, dependendo da arquitetura utilizada e do comprimento da sequência histórica. Ademais, a mudança no formato de representação da geolocalização deslocou a posição desse ponto de transição em questão do tamanho das sequências históricas.

No contexto deste estudo, ao utilizar a representação geográfica baseada em latitude e longitude (WGS84), a arquitetura Transformer e LSTM Entrelaçados (TLE) demonstrou desempenho superior em relação à LSTM e ao XGBoost para comprimentos de sequência histórica entre 1 e 5. No entanto, a arquitetura LSTM começou a apresentar melhor desempenho a partir de comprimentos de sequência entre 6 e 8. Isso indica um ponto de transição da arquitetura ótima quando o comprimento da sequência histórica aumenta de 5 para 6. Ao utilizar a representação geográfica Universal Transverse Mercator (UTM), esse ponto de transição entre as arquiteturas de melhor desempenho ocorre entre os comprimentos de sequência 4 e 5.

Portanto, entende-se que, ao selecionar a arquitetura mais adequada para a predição do próximo sensor externo de veículos roubados, tanto o comprimento da sequência histórica quanto o formato de representação geográfica devem ser considerados para alcançar resultados otimizados. Entende-se, também, que há indícios de que a arquitetura Transformer e LSTM Entrelaçados (TLE) é vantajosa para predições utilizando sequências curtas de entrada (abaixo de 6 pontos utilizando WGS84 e abaixo de 5 adotando UTM).

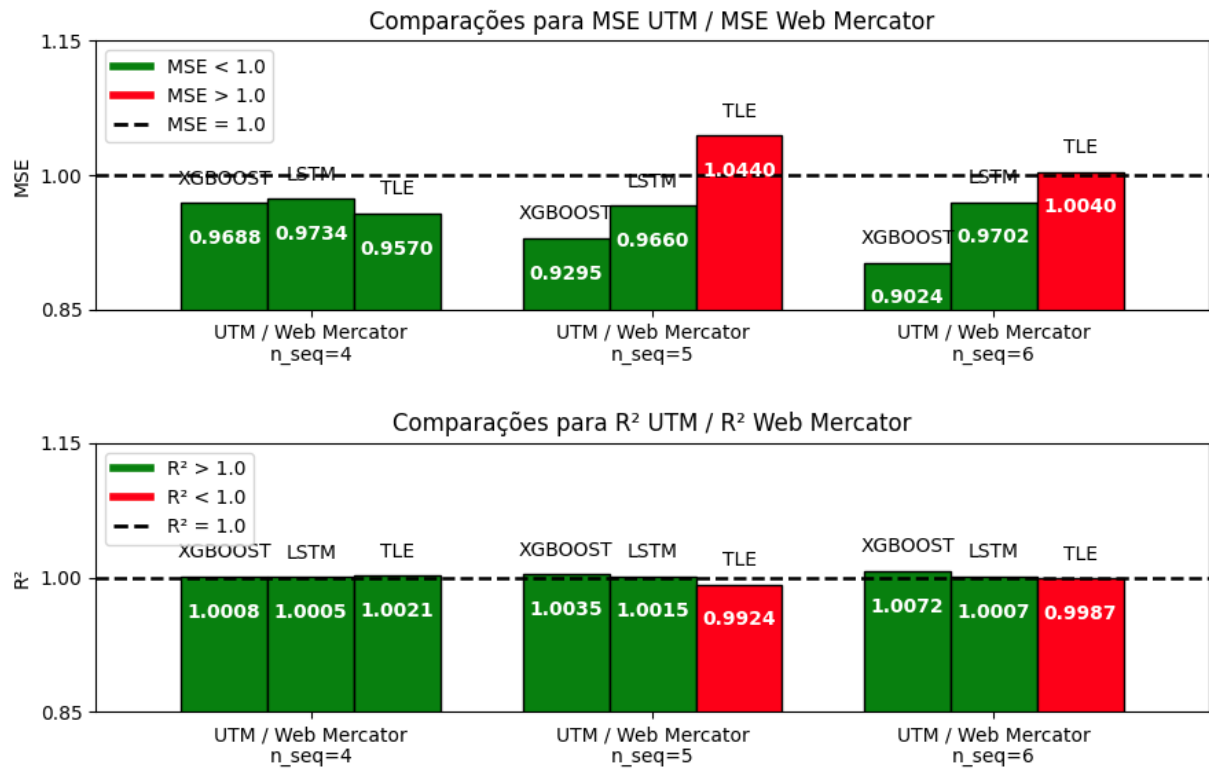


Figura 7.6: Comparação do formato de representação geográfica UTM sobre Web Mercator utilizando as métricas MSE e R².

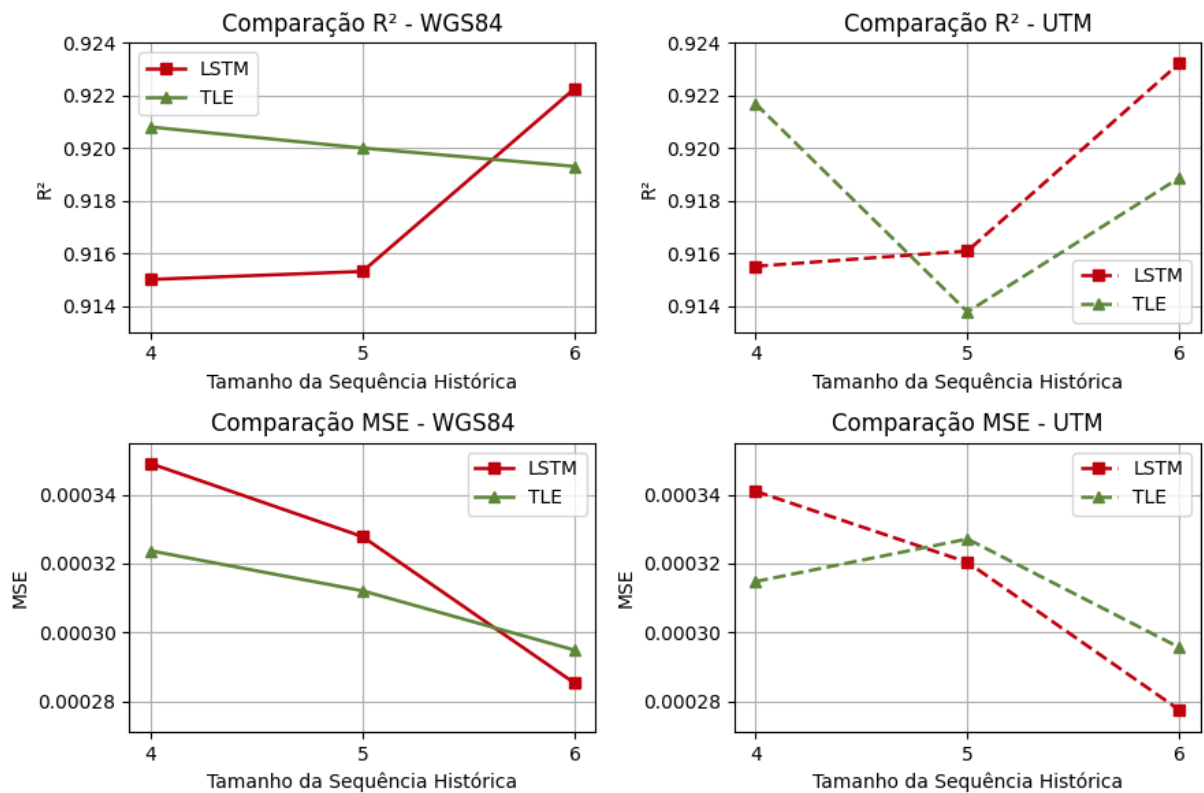


Figura 7.7: Ponto de transição das arquiteturas de melhor desempenho deslocado pela mudança de formato de representação geográfica.

8 CONCLUSÃO

Neste trabalho observou-se os desafios da predição de próximo sensor externo para trajetória veicular. Os dados esparsos, ainda que em volume considerável, tornam a predição desses veículos algo bem dificultoso, somando-se ao aspecto de provável comportamento anômalo desses motoristas em relação ao de um cidadão comum, tal predição torna-se um desafio ímpar. Para avançar o estado-da-arte nesse problema, foram planejados e executados dois estudos.

8.1 DISCUSSÕES FINAIS

O primeiro estudo, objetivando ganho na capacidade de predição, veio a incluir as informações sobre qual via da rua em que o veículo roubado foi identificado pelos sensores externos e, com um total de 54 cenários avaliados, obteve-se um ganho médio geral de 6,5% empregando a métrica de Erro Médio Quadrático (*Mean Squared Error* - MSE). Esse ganho médio foi calculado com base em ganhos em três tamanhos distintos de sequência histórica (tamanho 2 com 0,35% de ganho, tamanho 10 com 10% de ganho e tamanho 30 com 9,1% de ganho), demonstrando uma consistência em relação ao impacto positivo dessa informação nas predições desses veículos. Ainda nesse estudo, foi realizada a primeira análise da nova arquitetura Transformer e LSTM Entrelaçados (TLE) criada com o intuito de unir vantagens das arquiteturas de maior destaque na área de predição de sequências em *deep learning*, que são *Long Short-Term Memory* (LSTM) e Transformer. Essa nova arquitetura quando comparada às suas inspirações de maneira individualizada, apesar de não ter sido vantajosa para todos os cenários, apresentou um ganho de 0,74% na métrica MSE para o tamanho de sequência histórica de dois pontos de entrada, ficando a LSTM com o melhor desempenho para os tamanhos 10 e 30, também analisados.

Nesse primeiro estudo, então, atingiu-se o objetivo de ganho de capacidade de predição, tanto incluindo a informação da via da rua na predição, quanto para a sequência de dois pontos ao utilizar a nova arquitetura TLE. Concluiu-se, portanto, que a informação da via da rua traz pistas de para onde o veículo está indo, favorecendo a predição de próximo sensor externo de trajetória veicular para veículos roubados, mesmo em um contexto de comportamento anômalo do motorista. Concluiu-se, também, que a arquitetura Transformer e LSTM Entrelaçados (TLE) trouxe indicativos de ganho de predição para sequência histórica de dois pontos de entrada, ganho este importante, pois, no contexto da predição de próximo sensor externo para veículos roubados, quanto menor a quantidade de localizações históricas para predizer a próxima, melhor é para a sociedade, haja vista minimizar as chances de que o condutor do veículo roubado o utilize para cometer outros crimes subsequentes. Esses achados ainda identificaram a necessidade de uma análise mais aprofundada em relação a essa transição - entre as arquiteturas com melhor desempenho na predição de próximo sensor externo - relacionada ao tamanho da sequência histórica de entrada do modelo de Inteligência Artificial - IA.

Assim, observando a questão de transição entre arquiteturas, o segundo estudo foi executado com um total de 42 novos cenários, objetivando apoiar a definição de melhor arquitetura. Nesse estudo identificou-

se o exato ponto de transição entre as arquiteturas TLE e LSTM e sua dependência em relação ao tamanho da sequência histórica de entrada do modelo de IA. Além disso, observou-se que mudanças no formato de representação geográfica podem gerar um impacto positivo, negativo ou inconclusivo nas predições. A revelação de que o formato da representação geográfica das coordenadas de geolocalização dos veículos impacta tanto a capacidade de predição, quanto a posição do ponto exato de transição entre as arquiteturas foi surpreendente. Isso significa que, a escolha pela melhor arquitetura para predição de próximo sensor externo de veículos roubados deve ser posterior à definição do tamanho da sequência histórica e do formato da representação geográfica a serem armazenados e utilizados como dados de entrada do modelo para treinamento e predição.

No segundo estudo confirmou-se a vantajosidade da arquitetura Transformer e LSTM Entrelaçados (TLE) para sequências históricas curtas. Ainda, concluiu-se que, para a representação *World Geodetic System 1984* (WGS84; EPSG:4326), que utiliza latitude e longitude em graus, para sequências históricas iguais ou inferiores a 5 pontos (1 a 5 pontos), o ideal é aplicar a arquitetura TLE, e para sequências iguais ou superiores a 6 pontos (6 a 8 pontos), o ideal é aplicar a arquitetura *Long Short-Term Memory* (LSTM). No entanto, ao se empregar a representação geográfica *Universal Transverse Mercator* (UTM; EPSG:32723), que utiliza *Easting* e *Northing* em metros, deve-se atentar para a alteração do ponto de transição, onde para sequências históricas iguais ou inferiores a 4 pontos (1 a 4 pontos), idealmente aplica-se a arquitetura TLE, e para sequências iguais ou superiores a 5 pontos (5 a 8 pontos), idealmente aplica-se a arquitetura LSTM. Em relação ao formato Web Mercator (EPSG:3857), não se obteve ganhos consistentes, concluindo, portanto, que a conversão de um dos outros dois formatos - WGS84 ou UTM - para este não traria ganhos que compensem o processamento da conversão.

Em uma visão mais ampla, os resultados alcançados no contexto de veículos roubados podem fornecer subsídios em questão de direcionamento e otimização da aplicação de recursos públicos e estratégias de monitoramento das rodovias, assim como do emprego das forças policiais na questão da segurança pública. Nesse sentido, este trabalho trouxe um potencial ganho de eficiência para as instituições de segurança pública e contribuição para a recuperação dos veículos dos cidadãos, além da mitigação dos demais variados crimes subsequentes.

8.2 LIMITAÇÕES

Embora este trabalho contribua significativamente para a compreensão dessa nova arquitetura TLE, além do ganho de predição com a inclusão da informação sobre qual via da rua o veículo é identificado e de como o comprimento da sequência histórica e a representação geográfica afetam a escolha da arquitetura de predição mais adequada, algumas limitações devem ser observadas.

Uma primeira limitação a se observar se deve ao fato de o conjunto de dados ser restrito a registros de veículos roubados, o que pode limitar a generalização dos resultados para outros contextos de predição de trajetórias. Além disso, a análise baseia-se em sequências discretas e esparsas provenientes de sensores externos, um cenário que pode não capturar completamente os padrões complexos de movimento presentes em trajetórias contínuas.

Uma segunda limitação está relacionada ao ponto de transição das arquiteturas que, a depender da base de dados e em casos de outros formatos de representação geográfica, a transição pode ocorrer entre tamanhos distintos das sequências históricas identificadas neste trabalho. Essas considerações se aplicam também para o ganho da inclusão da informação sobre qual via da rua o veículo foi identificado, porém com menor impacto, pela abrangência dos resultados avaliados e por não se tratar de aspectos limítrofes como o ponto de transição, nem de apoio à definição de melhor arquitetura.

Outra limitação em relação a este trabalho é que não foram aplicadas técnicas de balanceamento dos dados, o que poderia impactar algumas análises. Entretanto, este é um ponto minimizado considerando a coerência dos resultados entre os estudos, mesmo utilizando quatro diferentes bases de dados, com dispersões distintas. Isso sugere robustez e boa generalização para os modelos ao lidar com a variação das distribuições, tornando essa limitação menos crítica e, paradoxalmente, um ponto positivo para a validação da abordagem adotada. As Figuras 8.1, 8.2, 8.3 e 8.4 apresentam, por meio de histogramas, as distribuições dos registros de latitudes presentes nas bases de dados do CE, GO, DF e SP, respectivamente; enquanto as Figuras 8.5, 8.6, 8.7 e 8.8 apresentam, também por meio de histogramas, as distribuições dos registros de longitudes encontradas nas bases de dados do CE, GO, DF e SP, respectivamente. Esse conjunto de Figuras de 8.1 a 8.8 demonstram essa diversidade de balanceamentos dos dados utilizados neste trabalho.

8.3 TRABALHOS FUTUROS

Para pesquisas futuras, restringir os períodos de tempo, realizando treinamentos e previsões para janelas de horários mais restritas (como período da manhã, tarde e noite) dos registros utilizados para treinamento e previsão pode aprimorar a precisão da previsão do próximo sensor externo, capturando padrões mais significativos no fluxo de veículos roubados em razão da variação dos aspectos comportamentais humanos ao longo do dia. Além disso, recomenda-se a expansão do conjunto de dados para incluir outros cenários de rastreamento veicular, além de veículos roubados, e a exploração da viabilidade de adaptar o modelo para trajetórias contínuas. Outra pesquisa interessante seria avaliar as interações de inclusão de novas informações que geram ganhos de desempenho, analisando se há sobreposição de ganhos. Em outras palavras, avaliar se os ganhos individuais de cada nova informação para treinamento e previsão se somariam totalmente ou se haveria uma interseção dos ganhos, gerando um ganho conjunto inferior à soma dos ganhos individuais. Essa análise poderia auxiliar a entender melhor os aspectos que mais contam para o aprendizado dos modelos no contexto de previsão de trajetória veicular.

Outro ponto interessante seria transformar o ID do sensor em uma categoria através de um *embedding* e concatenar com os atributos de coordenadas antes da camada densa de entrada do modelo. Além disso, poderia se realizar um teste estatístico, como Wilcoxon, para avaliar estatisticamente os resultados da capacidade da arquitetura TLE.

Essas melhorias podem aumentar a aplicabilidade do modelo e a precisão das previsões, permitindo sua implementação em uma gama mais ampla de contextos de monitoramento veicular, assim como utilizar a nova proposta de arquitetura Transformer e LSTM Entrelaçados (TLE) em contextos diferentes de previsão de sequências podem gerar resultados surpreendentes, especialmente para sequências curtas de entrada para

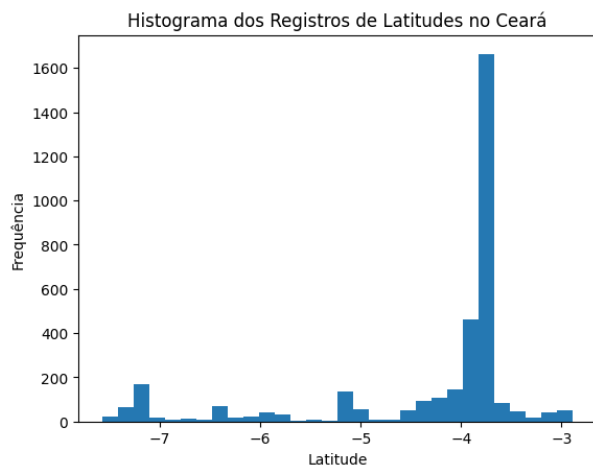


Figura 8.1: Distribuição dos registros de latitudes no Ceará (CE).

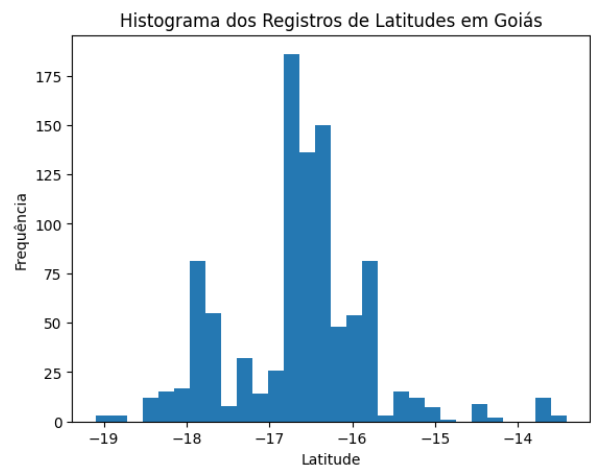


Figura 8.2: Distribuição dos registros de latitudes em Goiás (GO).

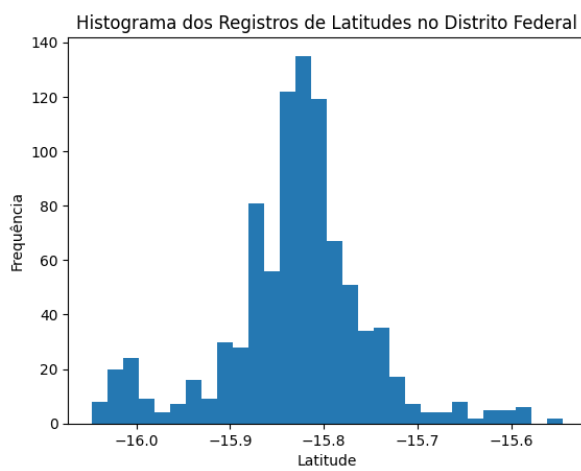


Figura 8.3: Distribuição dos registros de latitudes no Distrito Federal (DF).

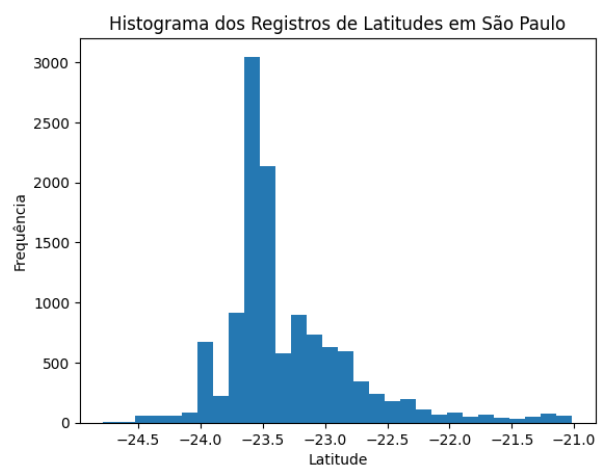


Figura 8.4: Distribuição dos registros de latitudes em São Paulo (SP).

o modelo de Inteligência Artificial - IA.

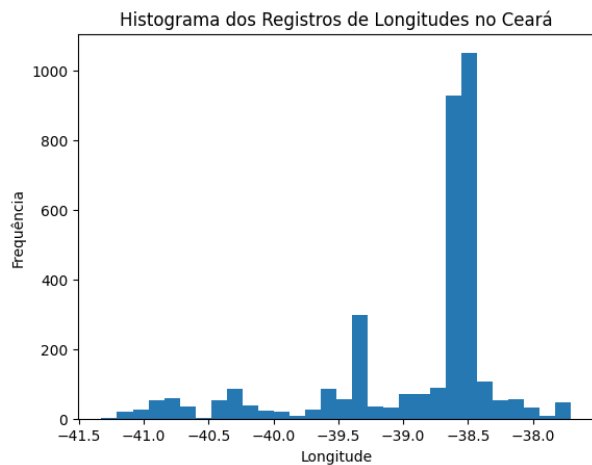


Figura 8.5: Distribuição dos registros de longitudes no Ceará (CE).

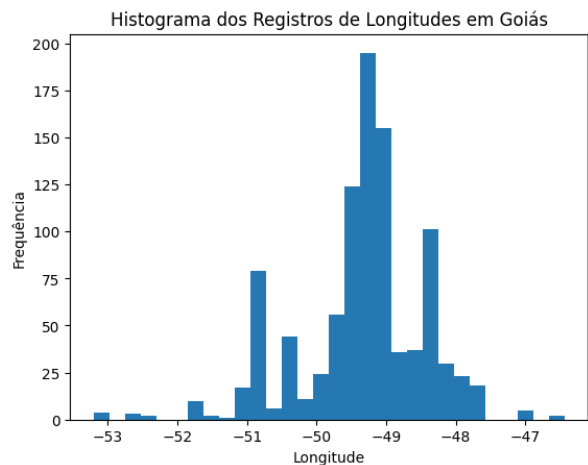


Figura 8.6: Distribuição dos registros de longitudes em Goiás (GO).

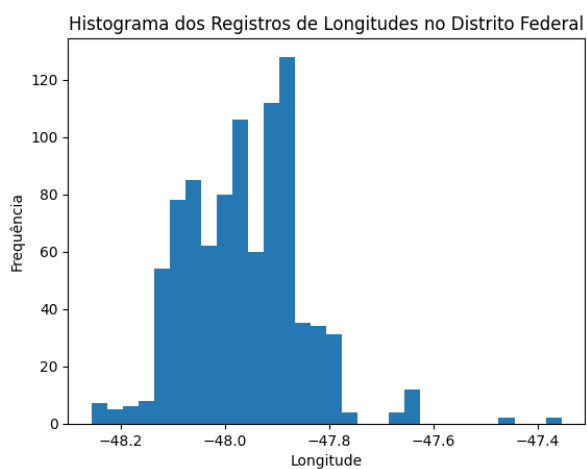


Figura 8.7: Distribuição dos registros de longitudes no Distrito Federal (DF).

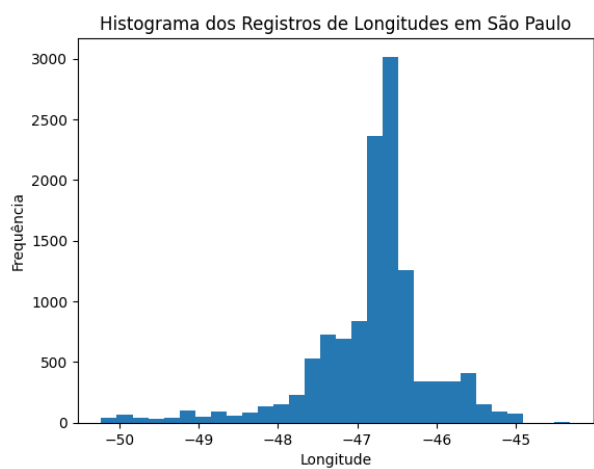


Figura 8.8: Distribuição dos registros de longitudes em São Paulo (SP).

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794.
- 2 TECH, D. *Como funciona a LSTM – Long Short Term Memory*. 2024. <<https://didatica.tech/lstm-long-short-term-memory/>>. [On-line; Acesso em: 17-12-2024].
- 3 VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017.
- 4 MATHIAS, J. M. A. C. *Coordenadas geográficas: latitude e longitude*. 2024. <<https://www.geogebra.org/m/un9hcccd>>. [On-line; Acesso em: 23-12-2024].
- 5 GURGEL, A. C. *A Projeção de Mercator*. 2024. <<http://www.historiadacartografia.com.br/projecao.html>>. [On-line; Acesso em: 23-12-2024].
- 6 GIOVANINI, A. *Projeção UTM: Características e Aplicações?* 2024. <<https://adenilsongiovannini.com.br/blog/projecao-utm-o-que-e-caracteristicas-e-aplicacoes/>>. [On-line; Acesso em: 23-12-2024].
- 7 PÚBLICA, F. B. de S. *18º Anuário Brasileiro de Segurança Pública: 2024*. [S.l.]: Fórum Brasileiro de Segurança Pública, 2024. <<https://publicacoes.forumseguranca.org.br/handle/123456789/253>>.
- 8 AUTOTEMPO, O. T. *Valor médio de carros usados deve superar os 80 mil no Brasil*. 2023. <<https://www.otempo.com.br/autotempo/valor-medio-de-carros-usados-deve-superar-os-r-80-mil-no-brasil-indica-estudo-1.3298557>>. [On-line; Acesso em: 23-12-2024].
- 9 FEDERAL, P. R. *SINAL - Sistema Nacional de Alarmes*. 2024. <<https://sinal.prf.gov.br/sicop/sinal>>. [On-line; Acesso em: 24-11-2024].
- 10 FEDERAL, P. R. *Alerta Brasil*. 2024. <<https://alertabrasil.prf.gov.br/alertabrasil/login>>. [On-line; Acesso em: 24-11-2024].
- 11 POPULAR, O. *Oito de dez veículos roubados são usados para prática de crimes | O Popular - Veja mais em: <https://opopular.com.br/cidades/oito-de-dez-veiculos-roubados-s-o-usados-para-pratica-de-crimes-1.88205>*. 2024. <<https://opopular.com.br/cidades/oito-de-dez-veiculos-roubados-s-o-usados-para-pratica-de-crimes-1.88205>>. [On-line; Acesso em: 24-11-2024].
- 12 BERNARDI, E.; MARTE, C. L.; YOSHIOKA, L. R.; RIBEIRO, P. C. M.; FONTANA, C. F. Modelo sistêmico e classificação de falhas associadas ao sistema de reconhecimento de placas para fiscalização automática de veículos. In: *Anais do XXIX Congresso de Pesquisa e Ensino em Transportes*. [S.l.: s.n.], 2015. p. 1510–1521.
- 13 NETO, J. S. D. S.; SILVA, T. L. C. D.; CRUZ, L. A.; LIRA, V. M. de; MACÊDO, J. A. F. de; MAGALHAES, R. P.; PERES, L. G. Predicting the next location for trajectories from stolen vehicles. In: *IEEE. 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*. [S.l.], 2021. p. 452–456.

- 14 HAVILAND, C. V.; WISEMAN, H. Criminals who drive. In: ASSOCIATION FOR THE ADVANCEMENT OF AUTOMOTIVE MEDICINE. *Proceedings: American Association for Automotive Medicine Annual Conference*. [S.l.], 1974. v. 18, p. 432–439.
- 15 TSILIGKARIDIS, A.; ZHANG, J.; TAGUCHI, H.; NIKOVSKI, D. Personalized destination prediction using transformers in a contextless data setting. In: IEEE. *2020 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2020. p. 1–7.
- 16 CHEN, J.; FENG, Q.; FAN, D. Vehicle trajectory prediction based on local dynamic graph spatiotemporal–long short-term memory model. *World Electric Vehicle Journal*, MDPI, v. 15, n. 1, p. 28, 2024.
- 17 WANG, W.; PENG, Z.; YANG, J. Single-sequence protein structure prediction using supervised transformer protein language models. *Nature Computational Science*, Nature Publishing Group US New York, v. 2, n. 12, p. 804–814, 2022.
- 18 YU, C.; MA, X.; REN, J.; ZHAO, H.; YI, S. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In: SPRINGER. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*. [S.l.], 2020. p. 507–523.
- 19 CRUZ, L. A.; SILVA, T. L. Coelho da; MAGALHÃES, R. P.; MELO, W. C. D.; CORDEIRO, M.; MACEDO, J. A. F. de; ZEITOUNI, K. Modeling trajectories obtained from external sensors for location prediction via nlp approaches. *Sensors*, MDPI, v. 22, n. 19, p. 7475, 2022.
- 20 HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural computation*, MIT press, v. 9, n. 8, p. 1735–1780, 1997.
- 21 GIULIARI, F.; HASAN, I.; CRISTANI, M.; GALASSO, F. Transformer networks for trajectory forecasting. In: IEEE. *2020 25th international conference on pattern recognition (ICPR)*. [S.l.], 2021. p. 10335–10342.
- 22 BAE, K.; LEE, S.; LEE, W. Transformer networks for trajectory classification. In: IEEE. *2022 IEEE International Conference on Big Data and Smart Computing (BigComp)*. [S.l.], 2022. p. 331–333.
- 23 GAIDUCHENKO, N. E.; GRITSYK, P. A.; MALASHKO, Y. I. Multi-step ballistic vehicle trajectory forecasting using deep learning models. In: IEEE. *2020 International Conference Engineering and Telecommunication (En&T)*. [S.l.], 2020. p. 1–6.
- 24 XU, Y.; WANG, Y.; PEETA, S. Leveraging transformer model to predict vehicle trajectories in congested urban traffic. *Transportation research record*, SAGE Publications Sage CA: Los Angeles, CA, v. 2677, n. 2, p. 898–909, 2023.
- 25 CHEN, X.; ZHANG, H.; ZHAO, F.; CAI, Y.; WANG, H.; YE, Q. Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for internet of vehicles. *IEEE Transactions on Instrumentation and Measurement*, IEEE, v. 71, p. 1–12, 2022.
- 26 MACEDO, G.; FILHO, G. R.; SANTOS, J.; NEVES, A.; ALMEIDA, M.; FALQUEIRO, M.; MENEGUETTE, R.; SERRANO, A.; MENDONÇA, F.; GONÇALVES, V. Predição de geolocalização de veículo com alerta de roubo usando lstm, transformer e tle. In: *Anais do XVI Simpósio Brasileiro de Computação Ubíqua e Pervasiva*. Porto Alegre, RS, Brasil: SBC, 2024. p. 61–70. ISSN 2595-6183. Disponível em: <<https://sol.sbc.org.br/index.php/sbcup/article/view/29403>>.
- 27 CRUZ, L. A.; ZEITOUNI, K.; MACEDO, J. A. F. de. Trajectory prediction from a mass of sparse and missing external sensor data. In: IEEE. *2019 20th IEEE International conference on mobile data management (MDM)*. [S.l.], 2019. p. 310–319.

- 28 CRUZ, L. A.; ZEITOUNI, K.; SILVA, T. L. C. da; MACEDO, J. A. F. de; SILVA, J. S. d. Location prediction: a deep spatiotemporal learning from external sensors data. *Distributed and Parallel Databases*, Springer, v. 39, p. 259–280, 2021.
- 29 (SBC), S. B. de C. *IV Seminário Grandes Desafios da Computação no Brasil – 2025-2035*. 2024. <<https://www2.sbc.org.br/grandesdesafios/computacao/>>. [On-line; Acesso em: 23-12-2024].
- 30 BRASIL, N. U. *Objetivo de Desenvolvimento Sustentável 16 Paz, Justiça e Instituições Eficazes*. 2024. <<https://brasil.un.org/pt-br/sdgs/16>>. [On-line; Acesso em: 23-12-2024].
- 31 HOLDSWORTH, M. S. J. *O que é deep learning?* 2024. <<https://www.ibm.com/br-pt/topics/deep-learning>>. [On-line; Acesso em: 23-12-2024].
- 32 AWAN, A. A. *O que é aprendizagem profunda? Um tutorial para iniciantes*. 2024. <<https://www.datacamp.com/pt/tutorial/tutorial-deep-learning-tutorial>>. [On-line; Acesso em: 23-12-2024].
- 33 BERGMANN, C. S. D. *O que é uma função de perda?* 2024. <<https://www.ibm.com/br-pt/think/topics/loss-function>>. [On-line; Acesso em: 23-12-2024].
- 34 ALAKE, R. *Explicação das funções de perda no aprendizado de máquina*. 2024. <<https://www.datacamp.com/pt/tutorial/loss-function-in-machine-learning>>. [On-line; Acesso em: 23-12-2024].
- 35 BOOK, D. L. *Capítulo 34 – O Problema da Dissipação do Gradiente*. 2022. <<https://www.deeplearningbook.com.br/o-problema-da-dissipacao-do-gradiente/>>. [On-line; Acesso em: 17-12-2024].
- 36 BOOK, D. L. *Capítulo 53 – Matemática na GRU, Dissipação e Clipping do Gradiente*. 2022. <<https://www.deeplearningbook.com.br/matematica-na-gru-dissipacao-e-clipping-do-gradiente/>>. [On-line; Acesso em: 17-12-2024].
- 37 OLAH, C. *Understanding LSTM Networks*. 2015. <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. [On-line; Acesso em: 17-12-2024].
- 38 BAHDANAU, D. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- 39 WIKIPEDIA. *EPSG Geodetic Parameter Dataset*. 2024. <https://en.wikipedia.org/wiki/EPSG_Geodetic_Parameter_Dataset>. [On-line; Acesso em: 24-11-2024].
- 40 IBGE. *Sistemas de Referência*. 2016. <https://geofpt.ibge.gov.br/informacoes_sobre_posicionamento_geodesico/sirgas/sisref_2.pdf>. [On-line; Acesso em: 24-11-2024].
- 41 WIKIPEDIA. *Web Mercator projection*. 2024. <https://en.wikipedia.org/wiki/Web_Mercator_projection>. [On-line; Acesso em: 24-11-2024].
- 42 GISGEOGRAPHY. *How Universal Transverse Mercator (UTM) Works*. 2016. <<https://gisgeography.com/utm-universal-transverse-mercator-projection/>>. [On-line; Acesso em: 24-11-2024].
- 43 GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. Disponível em: <<http://www.deeplearningbook.org>>.
- 44 BROWNLEE, J. *Regression Metrics for Machine Learning*. 2021. <<https://machinelearningmastery.com/regression-metrics-for-machine-learning/>>. [On-line; Acesso em: 24-11-2024].
- 45 BOTCHKAREV, A. Evaluating performance of regression machine learning models using multiple error metrics in azure machine learning studio. *Available at SSRN 3177507*, 2018.

- 46 ONOSE, E. *R Squared: Understanding the Coefficient of Determination*. 2023. <<https://arize.com/blog-course/r-squared-understanding-the-coefficient-of-determination/>>. [On-line; Acesso em: 24-11-2024].
- 47 OLIVEIRA, C. *Métricas para Regressão: Entendendo as métricas R^2 , MAE, MAPE, MSE e RMSE*. 2021. <<https://medium.com/data-hackers/prevendo-n%C3%BAmoros-entendendo-m%C3%A9tricas-de-regress%C3%A3o-35545e011e70>>. [On-line; Acesso em: 24-11-2024].
- 48 DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- 49 HENDRYCKS, D.; GIMPEL, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.
- 50 JOSEPH, V. R. Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, Wiley Online Library, v. 15, n. 4, p. 531–538, 2022.
- 51 SUN, R. Optimization for deep learning: theory and algorithms. *arXiv preprint arXiv:1912.08957*, 2019.
- 52 LOSHCHILOV, I.; HUTTER, F. *Decoupled Weight Decay Regularization*. 2019. Disponível em: <<https://arxiv.org/abs/1711.05101>>.
- 53 RASCHKA, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*, 2018.
- 54 BROWNLEE, J. *Train-Test Split for Evaluating Machine Learning Algorithms*. 2020. <<https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/>>. [On-line; Acesso em: 24-11-2024].
- 55 LIMA, M. *6 passos Para Criar Seu Primeiro Projeto de Machine Learning*. 2022. <<https://www.insightlab.ufc.br/6-passos-para-criar-seu-primeiro-projeto-de-machine-learning/>>. [On-line; Acesso em: 24-11-2024].