



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO (CTC)
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Felipe Vieira Roque

**A Time-Series Modeling Framework for Spatially-Segmented Emergency Event
Prediction Using Multimodal Data**

Florianópolis
2026

Felipe Vieira Roque

**A Time-Series Modeling Framework for Spatially-Segmented Emergency Event
Prediction Using Multimodal Data**

Tese submetida ao Programa de Pós-Graduação
em Ciência da Computação da Universidade Fed-
eral de Santa Catarina para a obtenção do título de
Doutor em Ciência da Computação.
Supervisor: Prof. Mateus Grellert, Dr.

Florianópolis
2026

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Roque, Felipe Vieira
A Time-Series Modeling Framework for Spatially
Segmented Emergency Event Prediction Using Multimodal Data
/ Felipe Vieira Roque ; orientador, Mateus Grellert, 2026.
139 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Ciência da Computação, Florianópolis, 2026.

Inclui referências.

1. Ciência da Computação. 2. Eventos de Emergência. 3.
TFT. 4. LSTM. 5. Séries Temporais. I. Grellert, Mateus.
II. Universidade Federal de Santa Catarina. Programa de Pós
Graduação em Ciência da Computação. III. Título.

Felipe Vieira Roque

**A TIME-SERIES MODELING FRAMEWORK FOR SPATIALLY-SEGMENTED
EMERGENCY EVENT PREDICTION USING MULTIMODAL DATA**

Esta Tese foi julgada adequada para obtenção do Título de
Doutor em Ciência da Computação, e aprovada em 17/12/2025 em sua forma final
pelo Programa de Pós-Graduação em Ciência da Computação do Centro Tecnológico
(CTC) da Universidade Federal de Santa Catarina. Membros:

Prof. Rafael de Santiago, Dr.

Universidade Federal de Santa Catarina –
UFSC

Prof.^a Renata Galante, Dra.

Universidade Federal do Rio Grande Do
Sul – UFRGS

Prof. Tiago Weber, Dr

Universidade Federal do Rio Grande Do
Sul – UFRGS

Certificamos que esta é a versão final e original do trabalho.



Documento assinado digitalmente

MATEUS GRELLERT DA SILVA

Data: 20/01/2026 11:01:54-0300

CPF: ***.664.940-**

Verifique as assinaturas em <https://v.ufsc.br>

Prof. Mateus Grellert, Dr.

Supervisor:

Universidade Federal de Santa
Catarina – UFSC



Documento assinado digitalmente

Carina Friedrich Dorneles

Data: 02/02/2026 17:09:21-0300

CPF: ***.834.250-**

Verifique as assinaturas em <https://v.ufsc.br>

**Programa de Pós-Graduação em
Ciência da Computação**

Florianópolis, 2026

To my wife, Jaqueline, whose steady and loving presence made every step of this journey possible. For the countless days of support, for the patience during my long absences, and for the resilience to understand the magnitude of this challenge. Your companionship was my anchor and my inspiration.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my family, especially my mother Dalva and my father Gilmar, whose unwavering support, values, and strength have guided me through every stage of my life. To my wife, Jaqueline, thank you for your patience, encouragement, and understanding throughout the many demanding moments of this journey. Your love has been fundamental to the completion of this work.

I also extend my sincere appreciation to my colleagues at CBMSC, whose collaboration and support were essential. In particular, I acknowledge Cap BM Sarraff for the assistance provided during the initial phase of this project, especially in facilitating access to data and offering continuous support.

My heartfelt thanks go as well to the professors who shaped my academic path. I am profoundly grateful to Professor Mateus Grellert, for whom words are insufficient to convey my appreciation. His countless hours of guidance, technical, scientific, and personal, were indispensable to my development. I also thank Professor Antonio Augusto Fröhlich, who guided me at the beginning of this journey and helped establish the foundations on which this research was built.

To all of you, my sincere gratitude.

“An intellect which at a certain moment would know all forces that set nature in motion, . . . , for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes”
(Pierre Simon Laplace)

RESUMO

Com o aumento do número de emergências e a ausência de soluções eficientes para prevê-las, a tarefa de alocar recursos tornou-se cada vez mais difícil. Esse desafio é agravado pelo fato de que as emergências podem estar relacionadas a contextos espaciais e temporais, exigindo ferramentas e técnicas especializadas para explorar essas relações de forma eficiente. Esta tese aborda a seguinte questão central de pesquisa: Podem as técnicas de aprendizado de máquina e métodos estatísticos prever com precisão eventos de emergência em diferentes horizontes temporais, melhorando assim a tomada de decisão quando combinadas com fontes de dados multimodais? Para responder a essa pergunta, propomos um *framework* para predição de eventos de emergência chamado FORESEE (*Forecasting Of Regional Emergencies using Spatial and Temporal Event Evaluation*). A novidade deste *framework* reside na integração de modelagem regionalizada por meio de segmentação espacial, no tratamento de dados mistos (multimodais) para aprimorar o desempenho preditivo e na previsão multi-passo em diferentes horizontes temporais. Os experimentos demonstram que combinar segmentação espacial com arquiteturas de *deep learning* aprimora significativamente a previsão de eventos de emergência em diferentes horizontes temporais. Experimentos conduzidos com vários anos de dados reais do sul do Brasil mostram que modelos especializados superam consistentemente abordagens generalistas. Por exemplo, modelos LSTM segmentados espacialmente alcançaram uma redução de 17,8% em MAE, saindo de 1,57 para 1,29, confirmando que a regionalização melhora significativamente o poder preditivo. Nossos resultados demonstram que a consideração de múltiplas macro-regiões ao longo de sete anos de dados, aliada à incorporação de informações multimodais, como condições climáticas e calendário, contribui para uma modelagem mais robusta e contextualizada das ocorrências. Além disso, o *Temporal Fusion Transformer* (TFT) superou tanto modelos tradicionais (como ARIMA) quanto abordagens de *deep learning*, alcançando ganhos de até 1,28 MAE em relação aos métodos concorrentes. Os mecanismos de interpretabilidade do TFT, como camadas de atenção e redes de seleção de variáveis, fornecem percepções sobre como variáveis exógenas influenciam dinamicamente as previsões, oferecendo uma transparência crucial para adoção operacional. O *framework* também possibilita previsões robustas de múltiplos passos à frente, produzindo resultados precisos desde horizontes de curto prazo até 15 dias adiante, o que é essencial para alocação proativa de recursos, planejamento de equipes e mitigação de riscos. Esses achados confirmam que técnicas de aprendizado de máquina e métodos estatísticos podem, de fato, prever com precisão eventos de emergência em diferentes horizontes temporais e, quando combinadas com dados multimodais, aprimoram substancialmente a tomada de decisão.

Palavras-chave: Eventos de Emergência, TFT, LSTM, Séries Temporais.

ABSTRACT

With the increasing number of emergencies and the absence of efficient solutions to predict them, the task of allocating resources has become increasingly difficult. This challenge is further aggravated by the fact that emergencies may be linked to spatial and temporal contexts, requiring specialized tools and techniques to efficiently explore these relationships. This thesis addresses the following central research question: Can machine learning techniques and statistical methods accurately predict emergency events across different temporal horizons, thereby improving decision-making when combined with multimodal data sources? To answer this question, we propose a framework for emergency event prediction called FORESEE (Forecasting Of Regional Emergencies using Spatial and Temporal Event Evaluation). The novelty of this framework lies in the integration of regionalized modeling through spatial segmentation, the handling of multimodal data to enhance predictive performance, and multi-step forecasting across different temporal horizons. Our experiments show that combining spatial segmentation with deep learning architectures significantly improves emergency event prediction across various time horizons. Experiments conducted with several years of real-world data from southern Brazil demonstrate that specialized models consistently outperform generalist approaches. For example, spatially segmented LSTM models achieved a 17.8% reduction in MAE, decreasing from 1.57 to 1.29, confirming that regionalization substantially enhances predictive power. Our results also show that considering multiple macro-regions over seven years of data, together with the incorporation of multimodal information such as weather conditions and calendar features, contributes to a more robust and context-aware modeling of emergency occurrences. In addition, the Temporal Fusion Transformer (TFT) outperformed both traditional models (such as ARIMA) and deep learning approaches, achieving gains of up to 1.28 MAE compared to competing methods. The interpretability mechanisms of TFT, such as attention layers and variable selection networks, provide insights into how exogenous variables dynamically influence predictions, offering crucial transparency for operational adoption. The framework also enables robust multi-step forecasting, producing accurate results from short-term horizons up to 15 days ahead, which is essential for proactive resource allocation, team planning, and risk mitigation. These findings confirm that machine learning techniques and statistical methods can indeed accurately predict emergency events across different temporal horizons and, when combined with multimodal data, substantially enhance decision-making.

Keywords: Emergency, Events, TFT, LSTM, Time Series.

LIST OF FIGURES

Figure 1 – Decomposition of a time series and its components. Source: Author's elaboration.	25
Figure 2 – White Noise Time Series. Source: Developed by the author.	28
Figure 3 – Random Walk Time Series. Source: Developed by the author.	29
Figure 4 – AR(4) Time Series. Source: Developed by the author.	30
Figure 5 – Several Time Series forecasting techniques. Source: Developed by the author.	31
Figure 6 – Time series modeling-forecasting methodology in a Machine Learning approach. Source: Developed by the author.	33
Figure 7 – Internal structure of an <i>Long Short Term Memory Neural Network</i> (LSTM) cell. Based on (OLAH, 2015)	35
Figure 8 – Architecture overview of Temporal Fusion Transformer (TFT). Based on (LIM et al., 2021).	42
Figure 9 – Pipeline from call to resource allocation for the emergency. Adapted from (MUKHOPADHYAY et al., 2020)	48
Figure 10 – Map view of the CBMSC dispatcher system interface.	49
Figure 11 – Experiment blueprint. Source: (CORTEZ et al., 2018)	57
Figure 12 – Two LSTM Models considering independence between clusters. Source: (CORTEZ et al., 2018)	58
Figure 13 – Proposed framework.	64
Figure 14 – Methodology to segment the regions.	67
Figure 15 – Multimodal data and predictor modeling approach.	69
Figure 16 – Time Series and Data Analysis.	71
Figure 17 – Evaluation and Interpretability.	73
Figure 18 – Workflow of the FORESEE framework, showing the main pipeline (A→B→C→D→E) and the iterative loops between exploratory analysis, regional modeling, multimodal Data, and evaluation.	76
Figure 19 – Framework components applied in Case Study 1.	78
Figure 20 – A representation of occurrences by day of the week and year.	81
Figure 21 – ACF and PACF plots for pre-hospital occurrences (DS-ambulance).	82
Figure 22 – Decomposition of the series into trend and seasonal for all data.	83
Figure 23 – Residual Component.	85
Figure 25 – Best k value for the problem and its visual separation seem from silhouettes for DS-Truck.	86

Figure 24 – Visual illustration of the data preparation methodology.	86
Figure 26 – Four clusters for DS-Ambulance events over time divided by the clustering technique with moving average per quarter.	88
Figure 27 – Runs optimization for DS-Ambulance occurrences.	90
Figure 28 – Loss plot for Ambulance occurrences.	91
Figure 29 – Error for different approaches.	95
Figure 30 – Framework components applied in Case Study 2.	98
Figure 31 – Daily occurrences for Macro-Flo.	100
Figure 32 – Weather variables.	101
Figure 33 – Daily Total Occurrences (2019) in Macro-Flo with Special Days. . .	103
Figure 34 – Spatial segmentation procedure to generate regions data for the models.	104
Figure 35 – Best k value for the problem and its visual separation seems from silhouettes for DS-Ambulance.	105
Figure 36 – Optimization results for DS-Ambulance occurrences.	105
Figure 37 – Attention in different scenarios.	112
Figure 38 – Encoder length in different scenarios.	113
Figure 39 – Decoder length in different scenarios.	114
Figure 40 – Predictions by cluster 15 days ahead to help decision making. . . .	115
Figure 41 – Past 60 days showing the regions that are dangerous to occur- rences in the next days.	116
Figure 42 – ACF and PACF plots for DS-Truck occurrences.	135
Figure 43 – Best k value for the problem and its visual separation seem from silhouettes for DS-Truck.	135
Figure 44 – Four clusters for DS-Truck events over time divided by the clustering technique with move average for 30-day.	136
Figure 45 – Best k value for the problem and its visual separation seem from silhouettes for DS-Truck.	137
Figure 46 – Four clusters for DS-Truck events over time divided by the clustering technique with move average for 30-day.	138

LIST OF TABLES

Table 1 – Summary table with the main works and proposed solution (MM = multimode, MS = multistep, SS = spatial segmentation).	60
Table 2 – Real occurrence data.	65
Table 3 – Data result with cluster assignments and occurrence counts.	68
Table 4 – Multimodal integration with weather variables added to regionalized data.	70
Table 5 – Complete multimodal table with temporal and calendar features added.	71
Table 6 – Real occurrence data.	79
Table 7 – Data description.	80
Table 8 – Number of clusters per average value of silhouette.	87
Table 9 – Description of cluster data.	88
Table 10 – Pearson correlation of tested parameters with the Mean Absolute Error.	92
Table 11 – Results of Emergency Event's regression models.	93
Table 12 – Parameters optimization for different approaches.	93
Table 13 – Comparing lstm regression results for different look ahead scenarios.	94
Table 14 – LSTM error metrics for the city of Itajaí (DS-ITA).	96
Table 15 – Number of occurrences for each city across the studied macro-regions.	99
Table 16 – Parameters optimization for different approaches.	106
Table 17 – Results for different approaches.	108
Table 18 – Results for different $e/$ and $p/$ windows in DS-Ambulance and DS-Truck datasets.	109
Table 19 – Mean Absolute Error (MAE) for different macro-region scenarios using DS-Ambulance and DS-Truck datasets for 1 day $p/$	110
Table 20 – Removing data from the TFT model to test predictability.	111
Table 21 – Error Metrics related to actual predictions to different clusters	115
Table 22 – Comparing retrained LSTM regression results for different look ahead values (direct forecasting).	136
Table 23 – Comparing LSTM regression results for a different look ahead by feeding back the input (recursive forecasting).	137

LIST OF ABBREVIATIONS AND ACRONYMS

ACF	Autocorrelation Function
ANN	<i>Artificial Neural Network</i>
ARIMA	Autoregressive Integrated Moving Average
dr	dropout rate
FORESEE	Forecasting Of Regional Emergencies using Spatial and Temporal Event Evaluation
hcs	hidden continuous size
hs	hidden size
lr	learning rate
LSTM	<i>Long Short Term Memory Neural Network</i>
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Mean Square Error
nah	number of attention heads
nhl	number of hidden layers
PACF	Partial Autocorrelation Function
RNN	<i>Recurrent Neural Network</i>
SVM	Supported Vector Machine
TFT	Temporal Fusion Transformer

LIST OF SYMBOLS

t	Time stamp
p	Autoregressive Order
q	Moving Average Order
d	Degree of Differencing
Q	"query"
K	"key"
V	"value"
T	Prediction Time

CONTENTS

1	INTRODUCTION	17
1.1	MOTIVATION	18
1.2	OBJECTIVES	21
1.3	CODE AND DATA AVAILABILITY	22
1.4	TEXT STRUCTURE	22
2	BACKGROUND	23
2.1	TIME SERIES	23
2.1.1	Modeling and forecasting Time Series	25
2.1.1.1	Traditional Approaches	27
2.1.1.2	Learning Based Models	32
2.1.1.2.1	<i>Long Short-Term Memory Neural Network - LSTM</i>	34
2.1.1.2.2	<i>Temporal Fusion Transformer</i>	36
2.1.1.3	Segmentation	42
2.1.1.3.1	<i>K-means</i>	43
2.1.1.3.2	<i>Sillouettes</i>	43
2.1.1.4	Performance Assessment Metrics	44
2.1.1.5	Model Validation	45
2.1.1.6	Multistep prediction	46
2.1.1.7	Hyperparameter Optimization	46
2.2	EMERGENCY RESPONSE	47
2.3	LIMITATIONS OF POPULATION-BASED STAFFING METRICS IN EMERGENCY SERVICES	49
2.4	MAIN CONCEPTS	51
3	STATE OF THE ART	53
3.1	EMERGENCY EVENTS	53
3.1.1	Urban Emergencies	53
3.1.1.1	Crime	54
3.1.1.2	Traffic Accidents	54
3.2	OVERVIEW OF THE STATE OF THE ART	55
3.3	MAIN RELATED WORK	56
3.3.1	<i>An architecture for emergency event prediction using LSTM recurrent neural networks</i>	56
3.4	SUMMARY AND RESEARCH OPPORTUNITIES	58
4	PROPOSED SOLUTION	62

4.1	THE FORESEE FRAMEWORK	63
4.1.1	Data Collection	65
4.1.2	Regionalization	66
4.1.3	Multimodal Data and Spatial-Temporal Modeling	68
4.1.4	Time Series and Data Analysis	71
4.1.5	Interpretability and Evaluation	72
4.1.6	How the FORESEE Should Be Used	74
5	CASE STUDIES	77
5.1	CASE STUDY 1: PREDICTION OF EMERGENCY EVENTS IN THE CITY OF FLORIANÓPOLIS	77
5.1.1	Data Collection	78
5.1.2	Time Series Analysis	80
5.1.2.1	Autocorrelation analysis	80
5.1.2.2	Times series decomposition	82
5.1.3	Regionalization	85
5.1.4	Spatio-Temporal Modeling	89
5.1.4.1	LSTM for predicting emergency events	89
5.1.4.2	LSTM performance comparison with different models	92
5.1.4.3	Comparing look-ahead LSTM scenarios	93
5.1.4.4	Model and framework generalization analysis	95
5.1.5	Discussion: A LSTM approach to predict emergency events using spatial features	96
5.2	CASE STUDY 2: TFT MODEL FOR MACRO-REGIONAL EXPERI- MENTS	97
5.2.1	Data Collection	98
5.2.1.1	Real MFC Occurrences	99
5.2.1.2	Weather Data	100
5.2.1.3	Calendar Days Data	102
5.2.2	Regionalization	102
5.2.3	TFT for emergency prediction	104
5.2.3.1	TFT for Predicting Emergency Events	104
5.2.3.2	Performance Comparison of Models	106
5.2.3.3	Multistep prediction with TFT model for different e_l	108
5.2.3.4	TFT in different Macro regions	109
5.2.4	Interpretability and Evaluation	110
5.2.4.1	Ablation analysis	110

5.2.4.2	Interpretability for different scenarios	111
5.3	CASE STUDY: REPRODUCING AN EMERGENCY SCENARIO WITH TFT MODEL	114
6	DISCUSSION	117
6.1	A FRAMEWORK PROPOSAL FOR EMERGENCY EVENTS PRE- DICTION	117
6.2	DISCUSSION OF THE FORESEE FRAMEWORK	118
6.3	ADDRESSING THE RESEARCH QUESTION	119
7	CONCLUSION	121
	references	123
	APPENDIX A – PUBLICATIONS SUMMARY	134
A.1	JOURNAL ARTICLES	134
A.2	CONFERENCE PAPERS AND OTHER PUBLICATIONS	134
	APPENDIX B – COMPLEMENTARY RESULTS	135
B.1	CASE STUDY 1	135
B.1.1	Additional vizualizations	135
B.1.2	Additional tables	136
B.2	CASE STUDY 2	137

1 INTRODUCTION

Time series refer to sequences of observations collected or recorded at successive points in time, often at uniform intervals (COWPERTWAIT; METCALFE, 2009). This form of data representation is commonly used to model a wide range of dynamic real-world phenomena, such as stock prices (CAVALCANTE, 2017), payroll metrics, product sales, urban temperatures, electricity consumption, and currency exchange rates. Due to their temporal dependencies and often complex structures, time series data present unique challenges in both modeling and prediction.

Forecasting future values using historical time series data is a well-established and essential research domain (AGGARWAL, 2015). In the literature on computational intelligence and statistical learning, time series forecasting is recognized as a non-trivial problem, as it requires extracting temporal patterns that can be influenced by trends, seasonality, abrupt changes, and noise. Traditionally, forecasting methods have been divided into two main categories: statistical models and data-driven approaches. Statistical techniques typically rely on predefined parametric assumptions about the underlying generative process (KUMAR; MURUGAN, 2013). In contrast, computational intelligence and machine learning models are inherently flexible and data-driven, enabling them to model both linear and non-linear dependencies without assuming a fixed functional form.

Emergency response management is a critical challenge that fire and emergency departments face worldwide. Emergency incidents like natural disasters (e.g., flooding, earthquakes), accidents (e.g., fires, explosions, traffic incidents), or public health crises are typically caused by a combination of multiple factors. To cope with the multidimensional and often non-linear characteristics of these events, researchers rely on statistical, analytical, and Machine Learning (ML) methods to develop data-centric emergency response mechanisms (HUANG, D. et al., 2021).

Managing these emergencies poses unique challenges, such as handling events with no established rules, limited information and resources (like ambulances and health professionals), and conflicting decision-making criteria (CHEN et al., 2019). This is further aggravated in densely populated regions, in which the frequency and complexity of emergencies have increased (RITCHIE; ROSER, 2018), partly due to increased traffic volumes and urbanization (MUKHOPADHYAY et al., 2020). Therefore, effectively predicting the occurrence of emergency events is key to optimizing resource allocation and facilitating effective communication across the emergency response network (JALDELL et al., 2014).

1.1 MOTIVATION

In many cities worldwide, population density is increasing (RITCHIE; ROSER, 2018). As a consequence, emergencies faced by rescue and relief teams have also risen. This can be partly explained by the fact that emergencies such as car accidents have grown in tandem with urbanization and traffic volume (MUKHOPADHYAY et al., 2020). Among the many challenges of managing emergencies, we can highlight the following: (1) events with no established rules; (2) a variety of constraints on time, information, and decision-making; (3) intangible and conflicting criteria (CHEN et al., 2019). Thus, organizations handling emergency resources must seek computational solutions to optimize the allocation and communication of the emergency response network (JALDELL et al., 2014). Furthermore, the aim of such resources should be to assist and make effective decisions with incomplete information during emergency response.

Developing effective Emergency Management (EM) systems and approaches has become increasingly important. The primary goal of such systems is to aid and enable emergency management in preparing for disasters and responding to urgent events (HUANG, D. et al., 2021). Emergency management systems have recently gained popularity, emphasizing the importance of learning from experiences to predict and prevent future emergencies by analyzing historical data from similar events or scenarios (CHEN et al., 2019). These facts confirm that various advanced technologies have been applied in EM systems to collect, process, and update spatial, temporal, and environmental information during emergency events, such as the Global Positioning System (GPS).

The availability of data and processing power has enabled various advancements in the theory and applicability of ML over the past decade. These areas hold promise for computing, analyzing, and predicting many events that impact people's lives (HADAD et al., 2019). The use of Machine Learning techniques can be found in different domains, such as education (WORSLEY; BLIKSTEIN, 2011), economics (HUANG, J. et al., 2020), and weather forecasting (XINGJIAN et al., 2015), among others. Regression models and statistical analysis can be useful for predicting emergencies (GREKOUSIS; LIU, 2019). More specifically, time can be used to model the relationship between incidents that can be statistically dependent (MUKHOPADHYAY et al., 2020).

Predicting emergency events locally and temporally can be challenging and has been addressed in previous works (CORTEZ et al., 2018; PETTET et al., 2017; VIEIRA

ROQUE et al., 2025; PAYARES-GARCIA et al., 2023; HUANG, D. et al., 2021). One of the most challenging issues in event prediction models is the varied characteristics of different spatial areas. This can be mitigated by creating region-specific models rather than generalized models, which lose the specificity of each location. Still, it also raises the question of how these regions can be defined automatically. A second problem is that building region-specific models can be time-consuming and unstandardized, making models more prone to errors and comparisons more difficult (VIEIRA ROQUE et al., 2025). Thus, solutions that automate this process are of utmost importance. Moreover, as highlighted by Di Huang et al. (2021), the development of real-time monitoring and prediction systems for emergency events remains an open problem in the literature, further underscoring the need for continued research in this area.

To mitigate these challenges, recent approaches have turned to ML techniques, particularly unsupervised learning and clustering algorithms, as a means of automating spatial segmentation (CORTEZ et al., 2018; VIEIRA ROQUE et al., 2025). These methods group emergency occurrences based on geographic coordinates, forming clusters that represent coherent spatial areas. Clustering enables the automatic identification of regions with similar emergency behavior, reducing manual intervention and improving reproducibility. To ensure meaningful and stable segmentation, internal evaluation metrics such as the Silhouette Score are commonly used to guide the selection of the optimal number of clusters.

Data variability is also a recurrent concern in contemporary modeling solutions. Information about a phenomenon or a system of interest can be obtained from different instruments, measurement techniques, experimental setups, and other sources. Due to the rich characteristics of natural processes and environments, a single acquisition method rarely provides a complete understanding thereof (LAHAT et al., 2015). In this sense, there is a need to handle these problems where one type of data is not enough to describe the behavior of a phenomenon, so many data sources must be merged.

In the context of emergency management systems, the nature of the data collected lends itself well to time series modeling. Emergency occurrences are typically recorded with precise timestamps, making it possible to construct temporal sequences that reflect the frequency and distribution of events over time. Many recent studies leverage ML and deep learning techniques to model these time series, aiming to capture patterns and dynamics solely based on historical occurrence data (MUKHOPADHYAY et al., 2020). While such models can learn temporal dependencies, their predictive power can be significantly enhanced by incorporating additional contextual information. By integrating multimodal features such as weather conditions,

traffic patterns, or public event calendars, the model gains a richer understanding of external factors that influence emergency dynamics (VIEIRA ROQUE et al., 2025).

Forecasts guide decisions in many areas of scientific, industrial, and economic activities, such as weather forecasting. Some problems require multi-step-ahead forecasting when the window of future events is for short, medium, or long horizons. Multi-step-ahead predictions consist of forecasting several future observations of a given sequence of historical observations (TAIEB, 2014). Many emergency event prediction approaches do not consider future varying windows for prediction. This limitation challenges effective decision-making, as emergency management often requires precise and adaptable forecasts to allocate resources efficiently and respond promptly to evolving situations.

The existing literature on emergency event prediction is often limited to specific regions or data types, restricting its generalizability and applicability, such as in (CORTES; VAPNIK, 1995). Previous studies typically focus on single data sources, which can result in incomplete understanding and suboptimal predictions (MUKHOPADHYAY et al., 2020). The research gap highlights the need for comprehensive approaches integrating diverse data sources. Our study addresses this by employing different occurrence data and a look-back analysis to inform the model's parameters. The main problem identified is the underutilization of multimodal data, which, if effectively integrated, can improve the precision and reliability of emergency event predictions.

This thesis advances the state of the art in emergency event prediction by proposing a comprehensive spatio-temporal multimodal forecasting framework designed to address key limitations in existing approaches. Emergency events are influenced by complex interactions between spatial distribution, temporal dynamics, and external contextual factors such as weather, mobility, and social activity. From a Computer Science perspective, this thesis contributes a structured and reproducible pathway for integrating heterogeneous data sources into spatial-temporal predictive modeling. This offers a new methodological direction for handling complex real-world datasets, where signals arise from different systems, resolutions, and contexts.

To overcome these challenges, we propose Forecasting Of Regional Emergencies using Spatial and Temporal Event Evaluation (FORESEE), an end-to-end framework that automates and structures the major stages required for accurate, explainable, and region-aware emergency event forecasting. The novelty of FORESEE lies in its integration of three core components:

- Regionalized modeling through spatial segmentation, enabling the creation of region-specific predictors that capture local behavioral patterns more effectively than generalized models;
- Multimodal data handling, which incorporates heterogeneous sources, such as climate, calendar events, and other contextual variables, to enhance the predictive capacity of the models;
- Multi-step forecasting enables predictions over various timeframes, accommodating both short- and medium-term forecasts.

1.2 OBJECTIVES

The research question defined for this work is as follows:

- Can machine learning and statistical techniques predict emergency events in different horizons, enhancing decision-making when integrated with multimodal data sources?

To address this question, the general objective of this study is to propose a comprehensive framework for training predictive models for emergency events using multimodal data types. This framework encompasses multimodal data, model training (utilizing both temporal and non-temporal techniques), and evaluation across various temporal windows. While the initial focus lies on data sourced from the Corpo de Bombeiros Militar de Santa Catarina (CBMSC), the framework is designed to be adaptable and extendable to other datasets on emergency event prediction, thereby facilitating broader applicability.

To achieve the defined objective, the following specific objectives have been identified:

- Review the literature on machine learning for emergency event prediction to identify research gaps;
- Analyze data from different sources (climate data, contextual data, etc) to enrich the analyses and model training;
- Evaluate external data to incorporate the data from real occurrences;
- Evaluate the data with time series analysis and descriptive statistical methods to detect patterns;

- Develop regionalization experiments that generate specialized predictive models for distinct spatial segments, enabling systematic comparison between regional and generalist modeling strategies;
- Assess the statistical relationships between multimodal variables and emergency occurrences;
- Evaluate multi-step prediction experiments capable of producing forecasts across multiple temporal horizons, from short-term intervals to extended windows;
- Investigate how different machine learning architectures perform under regionalized and multimodal data configurations for emergency event prediction.

1.3 CODE AND DATA AVAILABILITY

All the code and data used in the proposed framework are made available to the public to promote transparency and reproducibility of this research. The implementation of the framework, including data fusion, model training, and evaluation scripts, can be accessed on GitHub¹.

This open-source availability allows others to validate, utilize, and build upon the work presented in this thesis. Contributions and feedback are welcome to improve the framework continuously.

1.4 TEXT STRUCTURE

The remainder of the work is structured as follows. Chapter 2 addresses some fundamental theoretical concepts in the background structure. Chapter 3 presents the state of the art of the problem. The proposed Framework is outlined in chapter 4. The procedures adopted for conducting the emergency event experiment and its initial results are described in Chapter 5. Chapter 6 examines the primary outcomes associated with the framework and their implications. Chapter 7 outlines the findings and future work.

¹ <https://github.com/Felipe-Roque/OcurrencesPrediction>

2 BACKGROUND

This chapter presents a brief description of the methods and tools used to carry out the experiments. The experiments conducted and a brief description of the data can be accessed upon request¹.

2.1 TIME SERIES

Temporal data are common in data mining applications (AGGARWAL, 2015). This type of analysis results from some continuous or discrete event in which data is collected through hardware or software monitoring. Many applications are discrete time series, and those that are continuous can be discretized. Discrete sequences can be described as $\overline{X}_1 \dots \overline{X}_n$ of length n and dimensionality d , containing d discrete feature values at each of the n different timestamps $t_1 \dots t_n$. Each of the n components $\overline{X}_i$ contains d discrete behavioral attributes $(x_i^1 \dots x_i^d)$, collected at the i -th timestamp.

The definition, classification, and description of time series are activities related to identifying important components of a time series or its main behavior. The main components of a time series are trend, seasonality, noise, and outlier observations. A trend can be defined as a systematic change in a time series that does not appear to be periodic (COWPERTWAIT; METCALFE, 2009). It is an observable long-term feature, such as an increase or decrease in sequential data values. Trend analysis is a crucial task in time series analysis, as it provides insight into the overall or local behavior of the time series. Figure 1 presents the decomposition of a time series (second figure from the top down) showing the trend representation.

Seasonality is a repetitive pattern occurring at fixed or variable periods (COWPERTWAIT; METCALFE, 2009). Many real-world time series are affected by seasonal patterns, such as weather and sales time series, where a known pattern may repeat hourly, daily, weekly, monthly, yearly, etc. Seasonal analysis is another important task of time series analysis, as it helps to understand the periodic behaviors of the underlying process. Figure 1 illustrates a time series and the identification of seasonal patterns (third figure from the top down). Trends and seasonal components are often interpreted as deterministic components of a time series and modeled with mathematical functions. However, in more complex environments such as financial applications or emergency events, time series often exhibit random or stochastic trends. In other environments, these components may not be visible or exist.

¹ Available in: <https://github.com/Felipe-Roque/OcurrencesPrediction/blob/master/Cap2.ipynb>

Noise observations, outliers, and missing values are common in real-world time series. Noise and outliers are different observations from the aforementioned (trend and seasonality), and may be considered anomalous observations. It is worth noting that noise data in this context are observations of a time series and not the residual component resulting from trend and seasonality removal. The main difference between noise observations and outliers is that the underlying process does not genuinely generate noise, while outliers are genuine data points that may represent some peculiar situation in the time series processes. Missing values are unrecorded values within the fixed time interval of the time series. Noise observations and missing values typically originate from errors in the data collection process, such as meter failures or human errors. Both can hinder time series analysis and are often disregarded. Unlike noise, outliers may be particularly interesting and should not be excluded from analysis (AGGARWAL, 2015).

Depending on the behavior of the time series, it can be classified as linear or nonlinear, stationary or non-stationary. Linear processes interpret all regular structures in a dataset through linear correlations. This means that the intrinsic dynamics of the system are governed by the linear paradigm, where small causes lead to small effects. All irregular behaviors of the system are attributed to random factors. Therefore, linear time series are generated by a linear combination of present and past values and a random factor, and summarized by a linear model, such as a set of ordinary differential equations (BOX et al., 2015). Nonlinear time series, on the other hand, are those in which the observations of the time series result from a nonlinear combination of past values and random factors. In the case of a nonlinear system analysis, there are no universally applicable tools or methodologies.

A time series is considered stationary if its statistical properties remain the same throughout the time series observation process. The underlying stochastic process is in a state of statistical equilibrium. In practical terms, the mean and variance of time series observations are constant for x_t and x_{t-k} , for any lag k (FULLER, 1976). This type of time series shows no trend or seasonal patterns. Non-stationary time series, on the other hand, are those with different statistical properties throughout the generation process. This type of time series is more common in areas such as engineering, business, and economics. It is important to note that there are some ways to estimate and remove certain aspects of non-stationarity from a time series, such as trends and seasonality, to facilitate analysis. The simplest way to smooth out the non-stationary effects of trends and seasonality in time series is through differencing between successive data values x_t and x_{t+1} .

An important characteristic in some nonlinear time series data generation processes is the chaotic property. Chaos is a behavior of highly sensitive dynamic systems to initial conditions. Over time, chaotic systems can produce unpredictable, divergent, and infinitely detailed and self-similar behaviors (never truly repeating) due to this sensitivity (BOEING, 2016). Observations from a chaotic time series may randomly repeat several times without maintaining a defined periodicity (PALIT; POPOVIC, 2006).

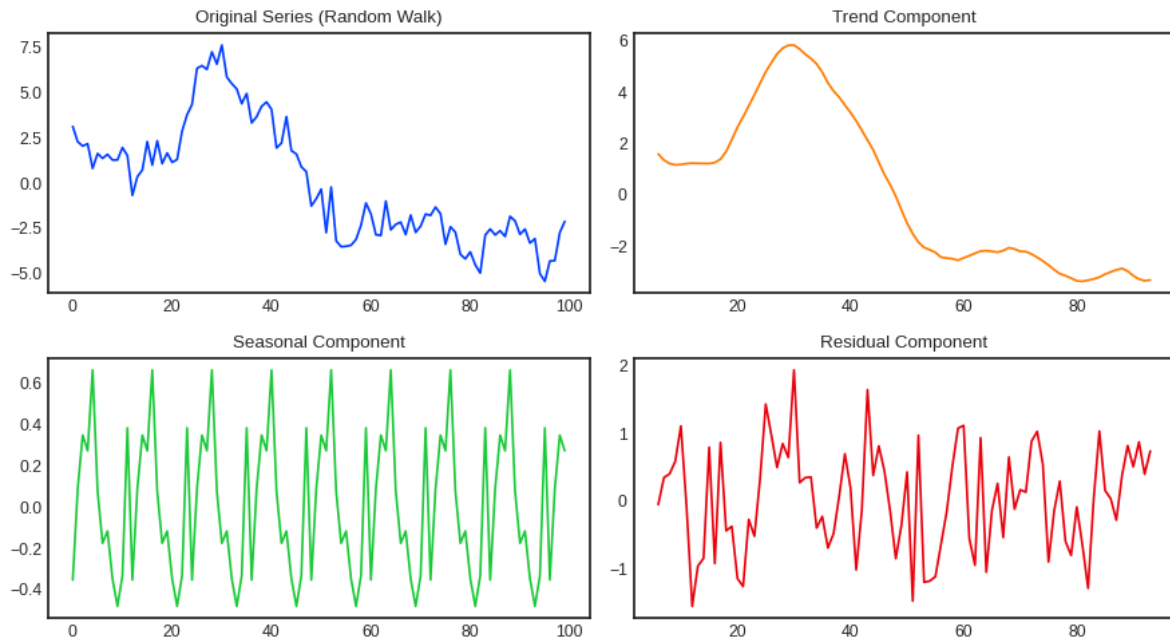


Figure 1 – Decomposition of a time series and its components. Source: Author's elaboration.

2.1.1 Modeling and forecasting Time Series

Building a model can serve as a means to provide a compact description of data or allow for predictions and simulations, which is considered one of the main tasks of time series analysis. A fitted model can help understand the past and forecast the future, providing informed decisions for managers or policymakers (COWPERTWAIT; METCALFE, 2009). Developing mathematical models capable of describing sampled data is one of the primary goals of time series analysis. The fitted model summarizes the main features of a time series.

Estimating a probability model for a time series differs from conventional data modeling due to a unique feature: the serial correlation of adjacent points in time (SHUMWAY et al., 2000). This serial correlation restricts the applicability of many

conventional statistical methods that depend on the main assumption that data observations are independent and identically distributed (CRYER, 1986). Therefore, the statistical models used to describe time series need to incorporate serial dependence to be suitable for predicting future values. Other features arise from this serial dependence that need to be addressed by the time series modeling process, such as the presence of trends and seasonality, stationary or non-stationary behaviors, linearity and non-linearity, and the presence of outliers, among others (CAVALCANTE, 2017).

Statistics and machine learning researchers have explored time series modeling due to its wide applicability. Statisticians seek to explain the behavior of time series through mathematical relationships between its components. Machine learning approaches aim to automatically extract the relationships that define time series behavior, without prior statistical assumptions about the data. According to Shumway et al. (2000), traditional statistical models of time series can be categorized into two main groups:

1. Time-domain models that seek to explain the behavior of the time series as a parametric function of current and past values;
2. Frequency-domain models that assume that the main feature of interest in time series are periodic or sinusoidal variations.

Although time-domain approaches are mainly used in engineering practice, these two approaches are not mutually exclusive and can be combined to improve modeling (PALIT; POPOVIC, 2006).

Decomposing the time series into deterministic and non-deterministic components is a common time-domain modeling approach used in the literature. This approach is based on quantifying these main components of the time series and the random variation or residuals (COWPERTWAIT; METCALFE, 2009). The deterministic components are trends and seasonal patterns, which can be modeled with mathematical time functions. When trend and seasonal components are modeled, they can be removed from the original time series, resulting in the random component. Then, another model can be fitted to the residuals, and forecasting can be achieved by predicting only the residuals. The forecasted value is then obtained by adding the residuals to the deterministic trend and seasonal models (BROCKWELL; DAVIS, 2002).

In these cases, simple differencing can serve to isolate the residuals from the time series, as this method does not depend on *a priori* assumptions about trends

and seasonal patterns, or even the assumption that these components remain fixed over the observation period (BROCKWELL; DAVIS, 2002). This scheme may not be properly used in time series that do not exhibit well-defined trends and seasonal patterns, such as emergency occurrence data, as in these cases, it is challenging to assume any deterministic component.

Box et al. (2015) proposed a general stepwise modeling strategy to guide the time series model estimation task. The three main steps of this process are (i) model specification or identification, (ii) model fitting, and (iii) model diagnosis. In the first step, a specific model needs to be chosen considering the characteristics of the time series to be modeled (the assumptions about the data). There are several families of statistical and computational intelligence models available in the literature. Examples of these main models are simple mean models, models considering trend and seasonality, linear models, complex nonlinear models, volatility models, frequency-domain models, and intelligent models (PALIT; POPOVIC, 2006).

To aid in identifying the most appropriate model, we can use some domain knowledge to be modeled, or we can perform an analysis of the time series characteristics through time series plots, scatter plots, and autocorrelation function, among other tools (COWPERTWAIT; METCALFE, 2009). An important requirement is that the chosen model has few parameters to be adjusted. Fitting several parameters may result in overfitting the model to historical data and consequently presenting a low extrapolation capacity.

After choosing the appropriate model (step (i)), the model parameters are estimated (ii). Model fitting consists of finding the best parameters to describe the observed data. Methods such as least squares and maximum likelihood are useful approaches for model fitting. Model diagnosis (iii) consists of evaluating the quality of the estimated model. Some accuracy metrics should be used to measure this quality. If any inadequacy is found in the constructed method, the process should return to the first step. The process must be restarted until an appropriate model is constructed, according to the defined objective criteria.

2.1.1.1 Traditional Approaches

Several mathematical models of time series have been proposed in the statistical literature. Each family of statistical models is designed to model specific time series behaviors and has its own peculiarities. The simplest model for a time series is the white noise model, designed for time series without trends and seasonal

components. In this model, observations are independent and identically distributed random variables with zero mean and finite variance. Time series generated from uncorrelated variables can be described as a Gaussian white noise model ωt , where $\omega t \sim N(0, \sigma_\omega^2)$. The stochastic behavior of a time series can be explained in terms of the white noise model, classical statistical methods are sufficient to describe that time series (SHUMWAY et al., 2000).

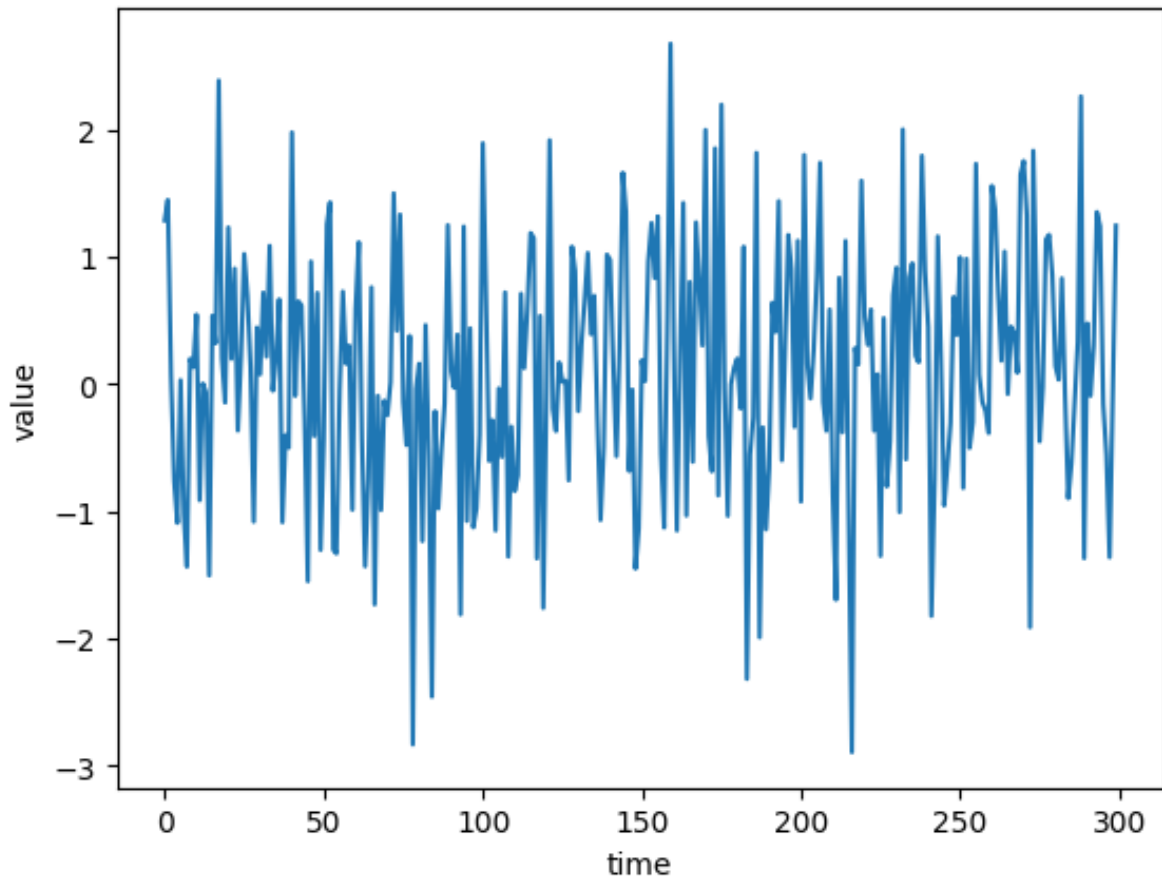


Figure 2 – White Noise Time Series. Source: Developed by the author.

A random walk generally provides a good fit for data with stochastic trends and no defined seasonal patterns (COWPERTWAIT; METCALFE, 2009), such as emergency event time series. In the increasing complexity scale of the model, the random walk model is obtained by cumulatively summing white noise observations. A zero-mean random walk is obtained by Equation 1, where ω_t is white noise. The behavior of the random walk with zero mean and with differentiation 1 can be observed in Figure 3. The Auto Correlation Function is plotted above to show the difference between the random walk with zero mean and with $diff = 1$.

When analyzing a time series using ACF, it often shows correlation at lag $L > 1$. However, this is typically due to the correlation at lag $L = 1$ that propagates to higher lags. Essentially, if the time series values at consecutive time points are correlated, this correlation can extend to subsequent time points, creating a chain of correlations that ACF captures as significant even at higher lags.

To isolate the direct correlation at each lag, we use the Partial Autocorrelation Function (PACF). The PACF measures the correlation between the time series and its lagged values, after removing the effect of shorter lags. In other words, PACF removes the influence of any intermediate lags, providing a clearer picture of the direct relationship at each specific lag.

To isolate the direct correlation at each lag we use the ACF with differencing (ACF with $\text{diff} = 1$). Differencing the time series removes the trend or systematic structure, helping to identify the true underlying correlations. We can also use the PACF function.

$$X_t = X_{t-1} + \omega_t \quad (1)$$

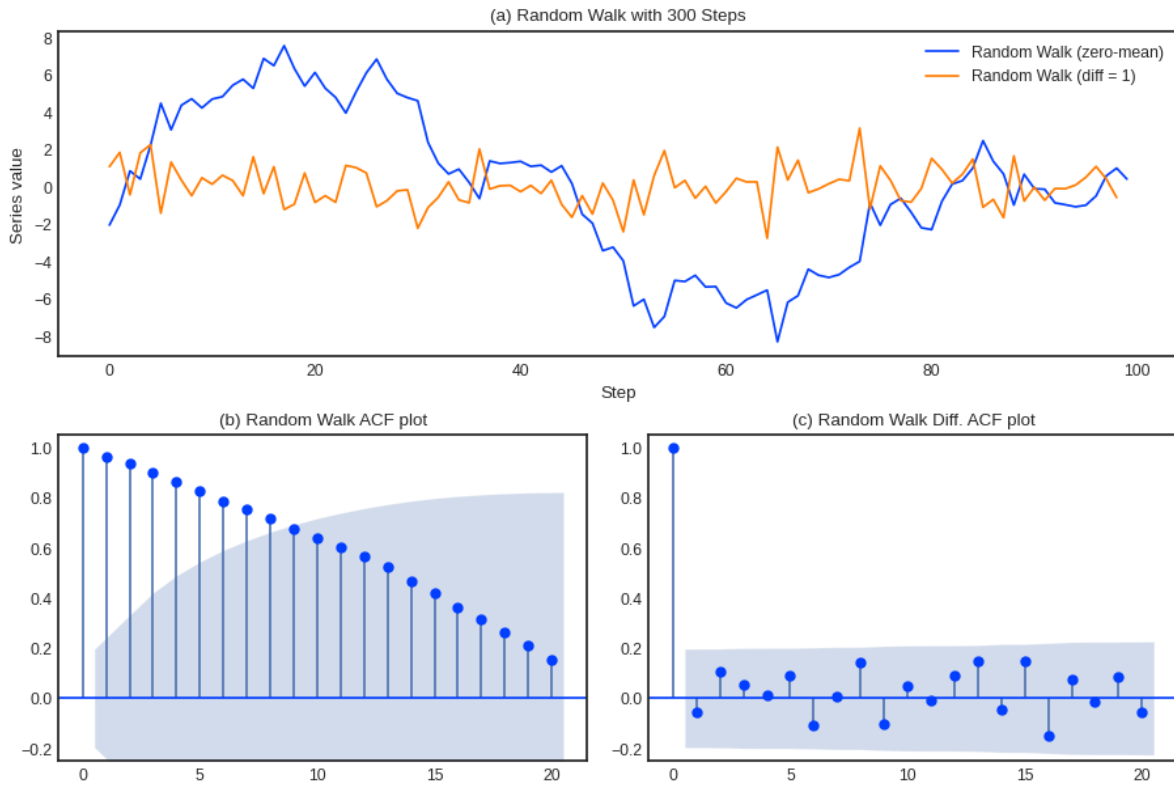


Figure 3 – Random Walk Time Series. Source: Developed by the author.

Another important linear model is the autoregressive (*AR*) model. It can be defined by a time series x_t following an autoregressive process of order p , known as $AR(p)$, as defined in Equation 2, where $\alpha_1, \dots, \alpha_p$ represent the model parameters and ω_t is white noise (COWPERTWAIT; METCALFE, 2009). A special case of the $AR(1)$ process is the random walk, where $\alpha_1 = 1$. Figure 5 shows the shape of a *AR* time series (in orange).

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \omega_t \quad (2)$$

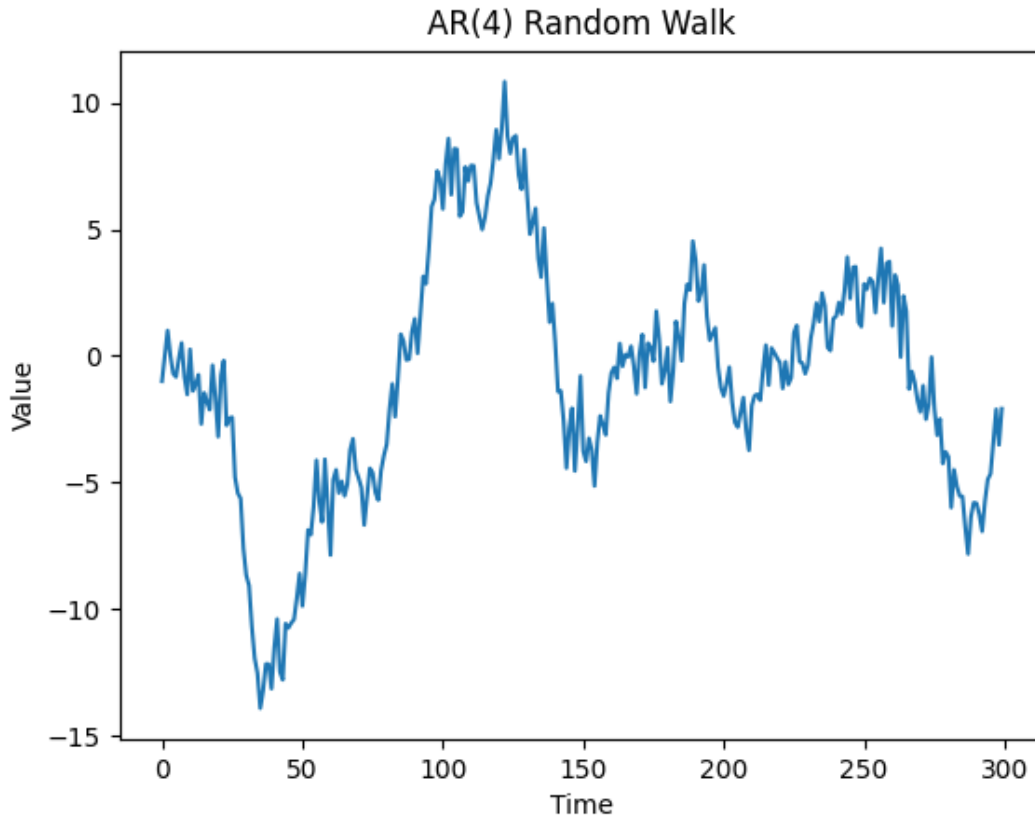


Figure 4 – AR(4) Time Series. Source: Developed by the author.

A suitable linear model for stationary time series is the moving average process. If a time series x_t can be expressed as a linear combination of the current white noise and the q most recent observations of white noise, it is called a moving average process ($MA(q)$) of order q , as defined in Equation 3. Figure 5 illustrates the shape of an $MA()$ time series.

$$x_t = \omega_t + \beta_1 \omega_{t-1} + \beta_2 \omega_{t-2} + \dots + \beta_q \omega_{t-q} \quad (3)$$

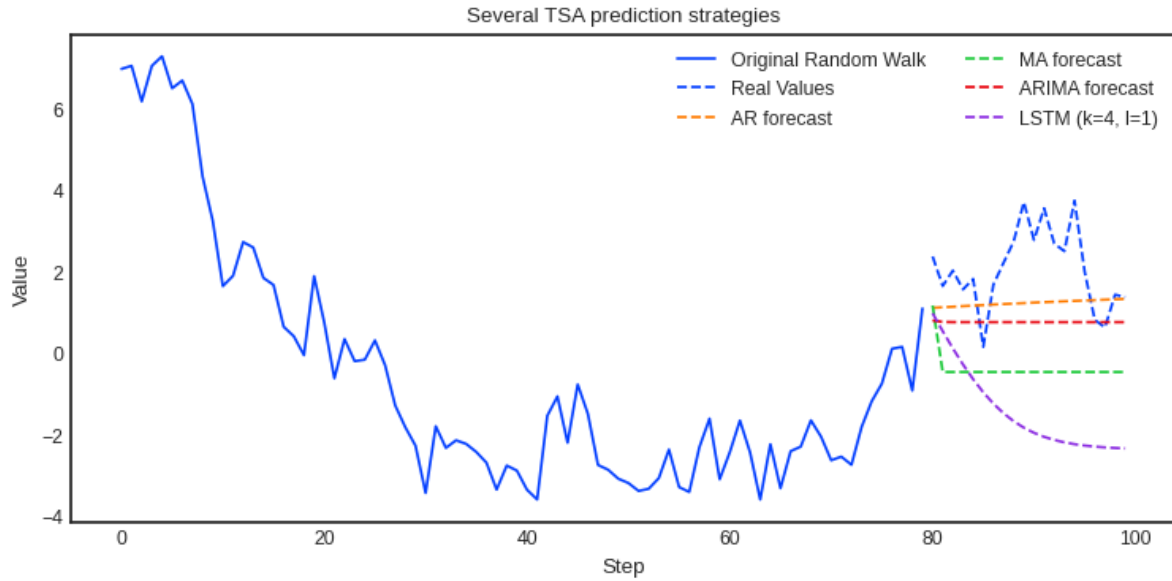


Figure 5 – Several Time Series forecasting techniques. Source: Developed by the author.

The Autoregressive Integrated Moving Average (ARIMA) model is a widely used class of linear models in time series modeling and forecasting. It combines autoregressive ($AR(p)$) and moving average ($MA(q)$) processes. A time series x_t follows an $ARIMA(p, d, q)$ process if the differences of order d of the series x_t form an $ARMA(p, q)$ process. In Figure 5, it is possible to visualize an $ARIMA(1, 1, 1)$ time series (in red).

Several other statistical models for Time Series have been proposed in the literature. Despite the extensive use of traditional statistical models for time series modeling and forecasting, some issues limit their application in real-world temporal applications. One such issue is the difficulty of automating the task of model specification (in the first phase of the overall model-building process). Choosing an appropriate model for a time series often involves analyzing time series plots and autocorrelations, assuming the behavior of time series, and requiring domain-specific knowledge. These tasks make automating time series modeling difficult. Another important issue is that traditional models often assume that the time series under study is generated from a linear process (KUMAR; MURUGAN, 2013). However, many time series, especially in economic areas, are inherently complex, highly noisy, dynamic, nonlinear, and chaotic

by nature (SI; YIN, 2013). Complex time series often require high-order polynomials to fully capture their behaviors. Moreover, many parameters can overfit the model to the data, resulting in inaccurate predictions.

2.1.1.2 Learning Based Models

In recent years, significant advances in applying computational intelligence methods in modeling and forecasting time series have occurred (LEE, 2009). These approaches stand out for their adaptability and flexibility, as they can capture both linear and nonlinear relationships among relevant factors without the need for prior knowledge about the input data (HUANG, D. et al., 2021). Among these techniques, *Artificial Neural Network* (ANN), *Supported Vector Machine* (SVM), and hybrid methods have been widely employed in time series forecasting, as they can estimate series behaviors without assuming prior statistical assumptions (GREKOUSIS; LIU, 2019; CORTEZ et al., 2018). Moreover, these methods exhibit high tolerance to inaccuracies and perform well in noisy environments, being numerical, data-driven, non-parametric, and self-adaptive mechanisms. Another advantage is that they require less historical data compared to traditional statistical models (CHENG; WEI, 2014). According to (CORTEZ et al., 2018), non-parametric methods are generally believed to outperform parametric methods in forecasting complex time series.

(PALIT; POPOVIC, 2006) outline a systematic procedure akin to that proposed by Box and Jenkins (1976) for forecasting time series using computational intelligence methods. This procedure encompasses the following operational steps: (i) data preparation, (ii) algorithm definition, (iii) training, and (iv) forecasting evaluation. Figure 6 summarizes these four steps in the machine learning modeling and forecasting methodology.

Data preparation marks the initial phase of time series modeling and forecasting. It begins with acquiring the data for learning and modeling purposes. Preprocessing procedures play a crucial role in enhancing the accuracy of intelligent predictors by selecting or extracting the most representative features, reducing input dimensionality, and minimizing training time. Techniques like *Principal Component Analysis* (PCA), *Independent Component Analysis* (ICA), *nonlinear ICA* (NICA), and *Wavelet Transform* (WT) are commonly employed for filtering noise, identifying outliers, and correcting them in the preprocessing stage (WANG, Jie; WANG, Jun, 2015).

Normalization involves scaling attribute values for training machine learning models. Data structuring entails organizing data for supervised learning, including

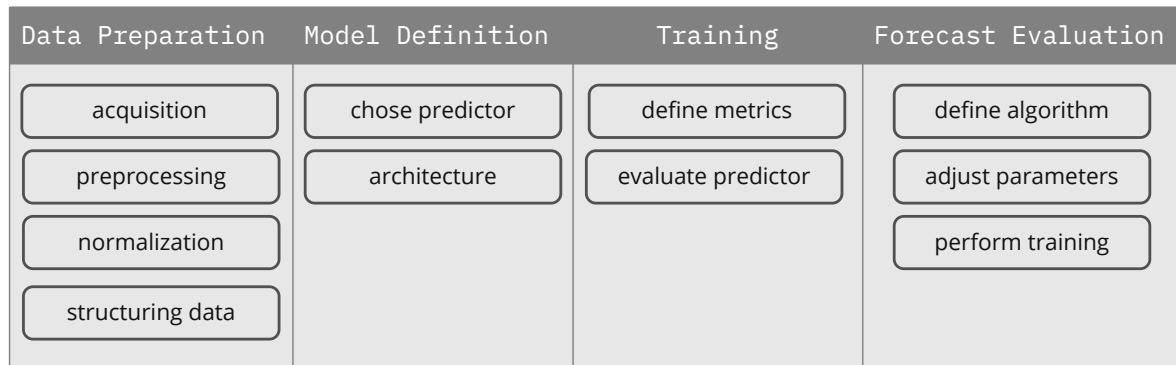


Figure 6 – Time series modeling-forecasting methodology in a Machine Learning approach. Source: Developed by the author.

partitioning data into training, validation, and test sets, defining input and output lags, and other relevant activities.

The model definition involves selecting the forecasting method and configuring its architecture. Various machine learning models, such as ANN, support vector regression, fuzzy systems, and ensemble techniques, among others, have been proposed for time series modeling and forecasting (PALIT; POPOVIC, 2006). Architecture configuration entails defining intrinsic model properties, such as the number of nodes, hidden layers, activation functions, and other architectural aspects, for models like Multilayer Perceptron (MLP).

In recent decades, various ANN models have been employed for forecasting future time series values. These include MLP trained with backpropagation ((ALKHEDER et al., 2017; YUAN et al., 2018), Functional Link Artificial Neural Network (FLANN) (MAJHI et al., 2009), LSTMs, (CORTES; VAPNIK, 1995), Radial Basis Function (RBF) Neural Network with Gaussian radial function, Bayesian Neural Networks (PETTET et al., 2017), among others.

SVM are computational intelligent learning methods often used as an alternative to ANN in pattern recognition tasks. SVM constructs a decision surface hyperplane to maximize the margin of separation between different class points, based on support vectors that define classification boundaries in the training set (Chen, 2010). Unlike empirical risk minimization, SVM implements the structural risk minimization principle to minimize an upper bound for the generalization error (Yuan, 2013). SVM's ability to produce global optimum solutions, compared to local optima generated by conventional ANNs, makes it valuable for both classification (SVC) and regression problems (SVR) (CHAMOLA et al., 2020).

After selecting the intelligent model, the next step is training it with available time series data. This phase involves choosing the training algorithm, adjusting parameters, and executing the training procedure to estimate a model of the learned data. Each training algorithm has specific parameters to adjust, such as the learning rate and the number of training epochs. For example, when using an MLP trained with backpropagation, the designer must define these parameters along with stopping criteria. Subsequently, the training process is carried out to fit the data.

Setting parameters for computational intelligence algorithms presents a significant challenge, as they can greatly influence forecasting accuracy. To address this issue, researchers have explored optimizing algorithms to find the best parameter settings and improve forecasting accuracy. Examples include Artificial Bee Colony (ABC), Genetic Algorithms (GA), and Particle Swarm Optimization (PSO), among others.

The final phase, evaluation, involves defining evaluation metrics and measuring the accuracy of results obtained by running the trained method on test data. It will be shown in Section 2.1.1.4.

2.1.1.2.1 Long Short-Term Memory Neural Network - LSTM

A *Recurrent Neural Network* (RNN) is a neural network with arbitrary connections between neurons. Typically, this connection is fully connected between adjacent layers. A standard RNN, with an input vector $v = (v_1, \dots, v_T)$ computes a hidden vector $h = (h_1, \dots, h_T)$ and an output vector $y = (y_1, \dots, y_T)$ by iterating equations 4 and 5 at $t = 1, \dots, T$.

$$h_t = \sigma(W_{ih}x_t + W_{hh}x_{t-1} + b_h) \quad (4)$$

$$y_t = W_{ho}h_t + b_o \quad (5)$$

where W_{ih} represents an input-to-hidden weight matrix, W_{hh} a hidden-to-hidden weight matrix, and W_{ho} an output-to-hidden weight matrix. b_h and b_o denote hidden and output bias vectors, and σ is usually a sigmoid function denoted by equation 6:

$$\sigma(x) = \frac{1}{1 + \exp(x)} \quad (6)$$

An LSTM can be defined as a RNN that can dynamically combine experiences due to its internal recurrences (MURTAGH et al., 2004). One of the major challenges of RNN is the vanishing/exploding gradient problem. LSTM is an implementation

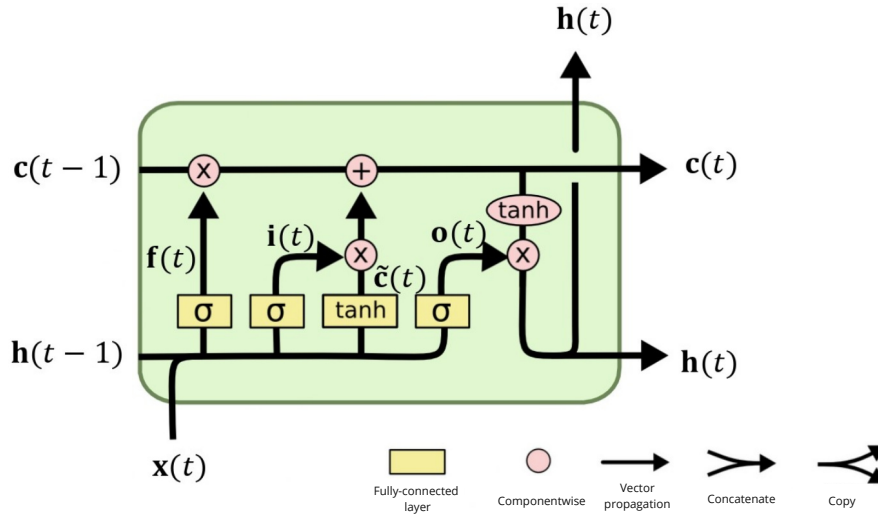


Figure 7 – Internal structure of an LSTM cell. Based on (OLAH, 2015)

of RNN that addresses the gradient problem. Additionally, LSTM cells can retain temporal series features in a context (HOCHREITER; SCHMIDHUBER, 1997). LSTM implements three gates, an input gate, an output gate, and a forget gate. These gates prevent the vanishing and exploding gradient problem by organizing the flow of information through the cells, Figure 7 shows an LSTM cell.

In Figure 7, $c(t)$ is the state vector that stores long-term information (or simply the long-term vector). As the network learns, useful information is stored in the state vector; otherwise, the information is discarded; $h(t)$ is the short-term vector, that is, the output of useful information from the cell (GRAVES; GRAVES, 2012).

Operations on the long-term vector are controlled by three gates: (i) the forget gate, (ii) the input gate, and (iii) the output gate. Depending on the chosen activation function, each element of the output vector takes on values in the range between 0 and 1. That is, when the value is 0, the information is removed, while for the value 1, the information is maintained. The following is a step-by-step description of the LSTM cell:

- Forget gate: Its importance lies in selecting the elements that must be stored at time t within the long-term vector or state vector, $c(t-1)$ (Figure 7), where $f(t)$ is the vector that controls the performance of the gate; $x(t)$ is the current input; $h(t-1)$ is the past output; as described by Graves (2012):

$$f(t) = \sigma([W_{hf}h(t-1) + W_{xf}x(t)] + b_f) \quad (7)$$

- Input gate: The purpose of this gate is to decide the intensity of the elements of the state vector to be modified. In this gate, $i(t)$ (Figure 7) represents the control signal of this gate, with a sigmoid activation function; in turn, $\tilde{c}(t)$ is a dense layer with a hyperbolic tangent activation function. In the dense layer, all new information should be stored in the state vector (GRAVES; GRAVES, 2012):

$$i(t) = \sigma([W_{hi}h(t-1) + W_{xi}x(t)] + b_i) \quad (8)$$

$$\tilde{c}(t) = \tanh([W_{hc}h(t-1) + W_{xc}x(t)] + b_c) \quad (9)$$

The new long-term vector $c(t)$ is composed of the addition of values stored by the forget gate and the new information from the input gate, that is:

$$c(t) = f(t) \odot c(t-1) + i(t) \odot \tilde{c}(t) \quad (10)$$

where $\odot$ represents the element-wise multiplication (Hadamard product).

- Output gate: The purpose of this gate is to choose which parts of the new long-term vector $c(t)$ should serve as the output of the LSTM cell. As illustrated in Figure 7, before serving as the output, the new long-term vector $c(t)$ passes through the tanh activation function.

$$o(t) = \sigma([W_{ho}h(t-1) + W_{xo}x(t)] + b_o) \quad (11)$$

$$\hat{y}(t) = h(t) = o(t) \odot \tanh(c(t)) \quad (12)$$

An LSTM cell neuron can be fed back by its input signal and by its output. Unlike a simple or deep feedforward network, which uses different parameters in each layer, the LSTM cell shares the same parameters at all steps. This reflects the fact that we are performing the same task at each step, just with different inputs. When dealing with time series, the feedback is accompanied by delays of one or more time units.

2.1.1.2.2 Temporal Fusion Transformer

The TFT, introduced by (LIM et al., 2021), is capable of incorporating a variety of static and dynamic input variables, making it a suitable architecture for complex

forecasting tasks such as emergency event prediction. The TFT uses several components to manage the complexity of multivariate time series forecasting (see 8). Let i represent unique regions in the emergency event time series. Each region i has a static covariate s_i . $x_{i,t}$ denotes other time-related inputs, and $y_{i,t}$ denotes the target variables at time t , $t \in [0, T_i]$.

Time-dependent inputs include two categories: $x_{i,t} = [o_{i,t}, k_{i,t}]$. Here, $o_{i,t}$ represents observed inputs, meaning they are unknown beforehand and can only be measured at time step t . $k_{i,t}$ represents known inputs, meaning they can be predetermined and their values are known prior to time step t .

This study uses observed inputs and targets up to time t , defined as:

$$\begin{aligned} y_{i,t-k:t} &= [y_{i,t-k}, \dots, y_{i,t}], \\ o_{i,t-k:t} &= [o_{i,t-k}, \dots, o_{i,t}], \\ k_{i,t-k:t+s} &= [k_{i,t-k}, \dots, k_{i,t+s}], \end{aligned}$$

where s is the prediction point. Finally, the prediction function is defined as follows:

$$\hat{y}_i(t, s) = f(s, y_{i,t-k:t}, o_{i,t-k:t}, k_{i,t-k:t+s}, s_i) \quad (13)$$

The TFT incorporates several innovative architectural components, enabling it to handle the complexity of emergency event prediction. TFT is designed to effectively process static, known, and observed input variables, ensuring robust predictions across diverse datasets and scenarios. The core components of the TFT architecture are as follows:

1. **Gating Mechanism:** This module selectively filters out irrelevant parts of the architecture, reducing computational overhead and enhancing model simplicity. This makes the TFT especially effective for large datasets and complex emergency prediction scenarios.
2. **Variable Selection Network:** By ranking input variables based on their importance for the target prediction, this component ensures that only the most critical features influence the model's decision-making process, ignoring less relevant inputs.
3. **Static Covariate Encoder:** This module encodes static features, such as regional or demographic information, which remain constant over time. Incorporating such information significantly improves the accuracy of predictions by providing context-specific insights.

4. **Temporal Dependency Module:** This module integrates a sequence-to-sequence layer and a multi-head attention mechanism. The sequence-to-sequence layer captures short-term temporal patterns essential for understanding recent emergency event trends, while the multi-head attention mechanism identifies long-term dependencies and interactions across the time series data.

In machine learning, increased model complexity does not always translate to improved predictive performance. To address this, the TFT employs the Gated Residual Network (GRN), designed to balance model complexity by adapting to input variability.

The GRN processes two input vectors: the primary information vector, denoted as p , and the context information vector, denoted as c . Its computation is governed by a series of operations, as outlined in Eqs. (14)–(18). The key feature of the GRN is its Gated Linear Unit (GLU), which acts as a gating mechanism to control the flow of information and suppress irrelevant components.

$$\text{GRN}_x(p, c) = \text{LayerNorm}(p + \text{GLU}_x(g_1)) \quad (14)$$

$$g_1 = W_{1,x}g_2 + b_{1,x} \quad (15)$$

$$g_2 = \text{ELU}(W_{2,x}p + W_{3,x}c + b_{2,x}) \quad (16)$$

The Exponential Linear Unit (ELU) activation function, employed in Eq. (16), is defined as:

$$f(x) = \begin{cases} x & x > 0 \\ e^x - 1 & x \leq 0 \end{cases} \quad (17)$$

The GLU, a pivotal component of the GRN, is responsible for the gate control mechanism. Given an input vector c , its formulation is as follows:

$$\text{GLU}_x(c) = \sigma(W_{4,x}c + b_{4,x}) \odot (W_{5,x}c + b_{5,x}) \quad (18)$$

Here, $\sigma(\cdot)$ represents the sigmoid activation function, $W(\cdot, \cdot)$ denotes weights, $b(\cdot, \cdot)$ represents biases, and $\odot$ indicates element-wise multiplication. By dynamically gating inputs, the GLU enables the GRN to streamline the model's structure, effectively ignoring unnecessary layers and focusing on relevant information.

In many real-world prediction tasks, the exact relationship between multiple input variables and the target variable is challenging to determine in advance. Consequently, it becomes difficult to gauge the specific contribution of each variable to the output. TFT addresses this challenge through its Variable Screening Network, which dynamically identifies the most relevant input variables and suppresses those that do not contribute meaningfully to prediction accuracy.

This module enhances the model's robustness by isolating critical variables and minimizing the influence of irrelevant or noisy data. Categorical variables are represented using entity embeddings, while continuous variables are transformed linearly. Each input variable is encoded as a d -dimensional vector, where d corresponds to the number of nodes in the network. Separate Variable Screening Networks are designed for different types of input variables, ensuring tailored processing for each input category. As the networks for static, known, and observed inputs share a common structure, this discussion focuses on the processing of past inputs.

The calculation process for determining variable screening weights is expressed in Eq. 19. Here, n_t represents the output of the GRN, and c_s is the context vector derived from the static covariate encoder. These weights are used to combine input variables, as illustrated in Eq. 20. Finally, as shown in Eq. 21, $\tilde{n}_t^{(i)}$ is computed by passing n_t through another GRN layer.

$$v_t^{(v)} = \text{softmax}(\text{GRN}_v(n_t, c_s)) \quad (19)$$

$$\tilde{n}_t = \sum_{i=1}^{m_x} v_t^{(v,i)} \cdot \tilde{n}_t^{(i)} \quad (20)$$

$$\tilde{n}_t^{(i)} = \text{GRN}_{\tilde{n}}(n_t^{(i)}) \quad (21)$$

Unlike many traditional time series prediction models, the TFT leverages static input variables. Static variables often represent fixed characteristics, such as geographic location or demographic data, which remain unchanged over time but can significantly influence the prediction outcome.

The TFT employs dedicated Gated Residual Network (GRN) encoders to process static inputs and generate four types of context vectors: c_s , c_e , c_c , and c_h . These context vectors play distinct roles within the Temporal Fusion Decoder, as outlined below:

1. Temporal Variable Selection (c_s): Used to prioritize and select relevant temporal variables dynamically.

2. Local Temporal Feature Processing (c_c, c_h): Facilitates localized processing of temporal features for short-term dependencies.
3. Temporal Feature Enrichment (c_e): Enhances temporal features by incorporating static information into their representation.

For instance, the temporal variable selection context vector c_s is derived as shown in Eq. 22:

$$c_s = \text{GRN}_{c_s}(f) \quad (22)$$

where f represents the static input features.

The TFT leverages a multi-head attention mechanism to capture long-term dependencies in time series data. This mechanism allows the model to focus on different parts of the input sequence simultaneously, enhancing its ability to identify intricate temporal correlations.

The self-attention mechanism, which forms the basis of multi-head attention, computes relationships between the Q , K , and V vectors. The calculation is given in Eq. 23:

$$\text{Attention}(Q, K, V) = A(Q, K) V \quad (23)$$

$$A(Q, K) = \text{Softmax} \left(\frac{QK^\top}{\sqrt{d_{\text{attn}}}} \right) \quad (24)$$

where d_{attn} is a scaling factor, ensuring numerical stability and appropriate weighting.

To enhance its representational power, the TFT employs a multi-head attention structure, which processes multiple attention mechanisms in parallel. This is defined as follows:

$$\text{MultiHead}(Q, K, V) = [H_1, \dots, H_m] W^H \quad (25)$$

$$H_h = \text{Attention}(QW_Q^{(h)}, KW_K^{(h)}, VW_V^{(h)}) \quad (26)$$

where m represents the number of attention heads, and $W_Q^{(h)}$, $W_K^{(h)}$, and $W_V^{(h)}$ are the learned projection matrices for each head.

To address the limitation that individual attention heads cannot share information, TFT aggregates the outputs of all heads additively. The improved multi-head attention mechanism is defined as:

$$\text{TFTMultiHead}(Q, K, V) = \hat{H} W \quad (27)$$

$$\hat{H} = \frac{1}{m} \sum_{h=1}^m A_h(Q, K) VW_V^{(h)} \quad (28)$$

This enhanced multi-head attention module allows the TFT to dynamically focus on various temporal features, significantly improving its ability to model complex patterns in time series data, such as those found in emergency event prediction scenarios.

To capture temporal relevance in the dataset, the Temporal Fusion Decoder (TFD) integrates multiple specialized layers, culminating in an output generation layer. The prediction step is denoted as $s \in \{1, \dots, s_{\max}\}$. The following layers are included in the TFD:

Time series data exhibit strong local dependencies, where the value at a given time step is influenced by nearby values. To model this, a sequence-to-sequence approach was employed to capture local dependencies. The input layer of the temporal fusion decoder uses gated skip connections, with its formulation expressed as:

$$\tilde{e}(t, n) = \text{LayerNorm}(e_n(t + n) + \text{GLU}_e(\phi(t, n))) \quad (29)$$

where e_n represents the local feature vector and ϕ denotes the gated input function.

Static variables significantly impact the accuracy of time series predictions. To incorporate this information, a static enrichment layer was introduced to encode the static input variables. The output is influenced by the position index n and the context vector c_e , as represented by:

$$\tilde{e}(t, n), c_e \quad (30)$$

The self-attention layer identifies the long-term temporal dependencies within the enriched features. At forecast time t , the static-enriched temporal feature matrix is denoted as $H(t)$. The matrix of attention outputs, $B(t)$, is defined as:

$$B(t) = \text{InterpretableMultiHead}(H(t), H(t), H(t)) \quad (31)$$

To facilitate effective training, an additional gating layer follows the self-attention mechanism.

This layer processes the outputs of the self-attention architecture and incorporates a GRN for enhanced connectivity to subsequent layers. The computation for this layer is defined as:

$$w(t, n) = \text{GRN}_w(d(t, n)) \quad (32)$$

$$\tilde{e}(t, n) = \text{LayerNorm}(\tilde{e}(t, n) + \text{GLU}_w(w(t, n))) \quad (33)$$

Finally, the prediction $\hat{y}$ at step s is generated using:

$$\hat{y} = W\tilde{e}(t, s) + b \quad (34)$$

where W and b are learned weights and biases, respectively.

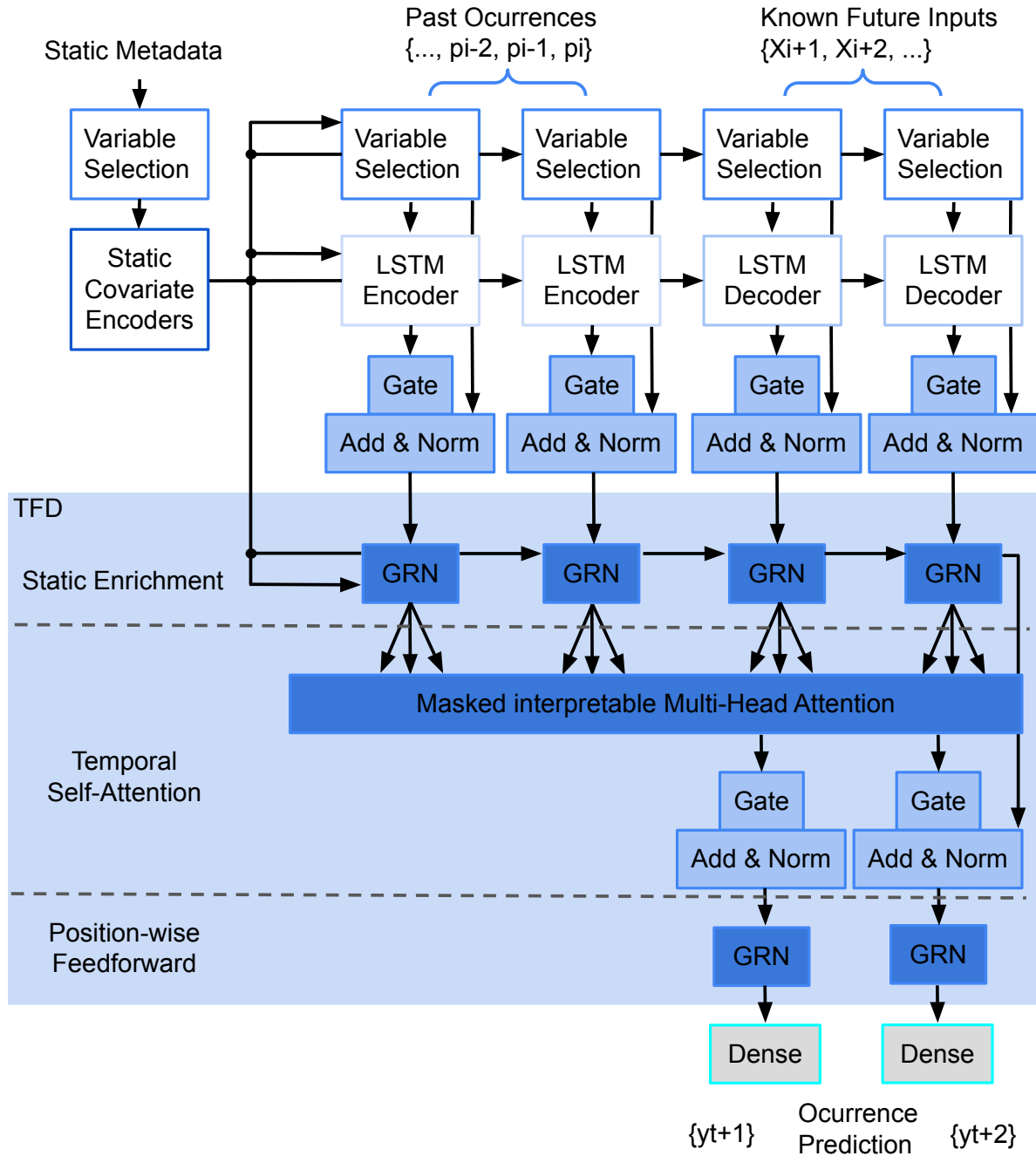


Figure 8 – Architecture overview of TFT. Based on (LIM et al., 2021).

2.1.1.3 Segmentation

Spatial clustering methods are classified into three categories: partition-based, hierarchical, and density-based algorithms (LI, X.; WANG, F., 2016). This study em-

employs three distinct spatial clustering algorithms. Specifically, K-means (LLOYD, 1982) and k-Medoids (RDUSSEEUN; KAUFMAN, 1987) are employed for partition-based algorithms, while Agglomerative (NIELSEN; NIELSEN, 2016) is selected for hierarchical-based algorithms.

2.1.1.3.1 *K-means*

The *K-means* algorithm groups examples into the same set based on their characteristics. In this way, the values of the centroids can be analyzed. Even so, in many cases, it is difficult, and on several occasions, it is not known how many groups the problem involves. The *Silhouettes* algorithm aids in this sense, assisting visually in identifying cases where the *K-means* is misplacing some examples.

K-means clustering (LLOYD, 1982) is an iterative partitioning algorithm that groups n observations into k non-overlapping clusters, where each observation is assigned to the cluster with the nearest centroid. The number of clusters k must be specified in advance. The algorithm seeks to minimize the within-cluster sum of squared distances between each observation and its corresponding centroid. In the context of emergency events, *K-means* has been employed to segment spatial data (VIEIRA ROQUE et al., 2025; CORTEZ et al., 2018), revealing distinct patterns and behaviors.

2.1.1.3.2 *Silhouettes*

To explore the data and analyze the groups created by K-means, the visual analysis algorithm Silhouette (ROUSSEEUW, 1987) was used. Silhouettes can help to find the best value of k for the problem since it is not easy to choose. The technique also visually assists in identifying cases in which k-means fail in some observations assigned to the clusters. One of the objectives of this approach is to segment localizations into different groups with well-defined attributes for a granular assessment of behaviors.

The value of Silhouette is defined by the equation 35 as follows:

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \quad (35)$$

$a(i)$ is the average distance between an example i and all other data within the same cluster. $s(i)$ can be interpreted as how well the given i object is inside his own cluster. $b(i)$ is the shortest average distance of i , for all points in any other cluster, of which i

is not part. Hence, the closer to 1 the value of $s(i)$, the more this observation belongs to the cluster. Similarly, negative values represent that this example does not belong to the cluster.

2.1.1.4 Performance Assessment Metrics

Forecasting evaluation typically employs regression error measures such as MAE, Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). These metrics are commonly used as figures of merit in time series forecasting studies, given by the Equations 36, 37 and 38, respectively.

The MAE represents the average value of the sum of absolute differences between outputs desired and outputs provided by the model:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (36)$$

The Mean Square Error (MSE) determines the mean of the squared errors, that is, the difference between average difference between the values predicted by the model and the actual values:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (37)$$

RMSE penalizes larger error terms and tends to become larger and larger than the MAE for outliers, which are obtained from the mean squared error:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}} \quad (38)$$

Y_i is the real value of the data and the $\hat{Y}_i$ is the predicted value of regression technique. n is the number of observations of the data set evaluated.

Akaike Information Criterion (AIC) (AKAIKE, 1969) is another important metric used for model selection, especially in time series analysis. The AIC provides a means for model comparison by balancing model fit and complexity, penalizing models with a larger number of parameters. The following equation gives it:

$$AIC = 2k - 2 \ln(L) \quad (39)$$

where k is the number of parameters in the model and L is the maximum value of the likelihood function for the model. The AIC is useful because it not only measures

the goodness of fit of the model but also includes a penalty term for the number of parameters, thus helping to avoid overfitting. A lower AIC value indicates a better model.

2.1.1.5 Model Validation

Validating a predictive model is a crucial step to ensure that its performance is reliable, generalizable, and not merely the result of overfitting to the training data. Traditional validation strategies, particularly k -fold cross-validation, are frequently used in machine learning due to their simplicity and effectiveness for static datasets (STONE, 1974). In standard k -fold validation, the dataset is randomly divided into k equally sized folds; the model is trained on $k - 1$ folds and evaluated on the remaining one. This process is repeated k times, and the average evaluation metric serves as an estimate of the model's expected performance.

However, while widely used, standard cross-validation cannot be applied directly to time-dependent data (ROBERTS et al., 2017). Randomly shuffling time-ordered records destroys the temporal structure that is fundamental to forecasting tasks. Such shuffling leads to data leakage, where the model gains access to future information during training, resulting in overly optimistic performance estimates that cannot be replicated in real-world scenarios.

To overcome this limitation, we employ Time Series Cross-Validation (TSCV), also known as *rolling-origin evaluation* or *walk-forward validation* (BERGMEIR; BENÍTEZ, 2012). In this strategy, the temporal order of the data is strictly preserved. The model is first trained on an initial window of observations and then tested on the immediately following period. The window is progressively expanded (or moved) forward in time, and the process is repeated across several folds.

Formally, given a time series

$$\{y_1, y_2, \dots, y_T\},$$

the splits generated by time series cross-validation follow the pattern:

$$\{y_1 \rightarrow y_{t_1}\} \text{ (train)} \rightarrow \{y_{t_1+1} \rightarrow y_{t_2}\} \text{ (test)},$$

$$\{y_1 \rightarrow y_{t_2}\} \text{ (train)} \rightarrow \{y_{t_2+1} \rightarrow y_{t_3}\} \text{ (test)},$$

and so on, until the end of the series.

This procedure ensures that each evaluation step respects causality, mimicking the real forecasting scenario in which only past data is available to predict the future.

2.1.1.6 Multistep prediction

There are two different methodologies to solve the problem of multistep time series prediction (STOCK et al., 2005). For a one-step forecasting problem, we can construct a training set X, Y of shape $[(N - n), n]$ and $[(N - n), 1]$ where N is the total length of the series, and n is the number of times we lag the series, often named window length. X represents a training example, while its target corresponds to the element in the same row in Y . This strategy only describes one step ahead of forecasting.

The first method, called the recursive approach, involves retraining the model with the same optimal parameters. As a result, each instance of the model is trained to predict the number of steps ahead. In other words, H models f_h are trained independently:

$$y_{t+h} = f_h(y_t, y_{t-1}, \dots, y_{t-n+1}) + w_{t+h} \quad (40)$$

with $t \in \{n, n + 1, \dots, N + H\}$ and $h \in \{1, 2, \dots, H\}$ where H is the forecast horizon. Every model outputs a prediction of one value. The predictions from the H models are then concatenated to form an H -step forecasting. This strategy has no accumulated error but requires a long computational time due to enormous models to learn, different from the second strategy that accumulates error for each new prediction.

The second strategy, called the direct method, is to use the previous model to predict multiple future time steps. This is done by recursively using the model's f output for subsequent predictions, effectively feeding the output back into the model as a look-back input.

$$y_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-n+1}) + w_{t+1} \quad (41)$$

with $t \in \{n, n + 1, \dots, N - 1\}$ and w_{t+1} is a noise term.

2.1.1.7 Hyperparameter Optimization

Hyperparameter optimization is a fundamental component of modern machine learning workflows, as the choice of learning rate, hidden size, number of layers, dropout rate, and other architectural parameters can significantly affect model accuracy, stability, and generalization. Instead of relying on manual tuning or exhaustive techniques such as grid search, we employ Optuna (AKIBA et al., 2019), a widely

adopted optimization framework designed to automate and accelerate the search for optimal hyperparameters.

Optuna follows a *define-by-run* paradigm, in which the search space is constructed dynamically during execution. This allows flexible specification of complex conditional spaces (e.g., choosing a different set of parameters depending on whether an LSTM or transformer-based model is selected). For each trial, Optuna samples a hyperparameter configuration, trains the model using that configuration, and evaluates an objective function, typically a validation error such as MAE.

Several optimization strategies are available within Optuna, including grid search, random search, and evolutionary algorithms. However, its default and most powerful method is the Tree-Structured Parzen Estimator (TPE) (BERGSTRA et al., 2011), a Bayesian optimization algorithm specifically designed to be more sample-efficient than classical approaches.

The TPE algorithm reformulates the optimization problem by modeling the conditional distribution of hyperparameters given the performance value y :

$$p(x | y) = \begin{cases} \ell(x), & y < y^* \\ g(x), & y \geq y^*, \end{cases}$$

where $\ell(x)$ represents good-performing configurations (those producing objective values better than a threshold y^*), and $g(x)$ models the poorer ones. The threshold y^* is typically a quantile (e.g., 20%) of previously observed objective values.

New hyperparameters are then chosen by maximizing the ratio:

$$x^* = \arg \max_x \frac{\ell(x)}{g(x)},$$

which prioritizes regions of the search space where high-performance configurations are more probable. This mechanism allows TPE to focus the search on promising hyperparameter areas while still exploring diverse possibilities.

2.2 EMERGENCY RESPONSE

Emergency incidents, such as an accident, for example, are usually reported to emergency response agencies (MUKHOPADHYAY et al., 2020). These agencies have mechanisms to process these calls. In Brazil, there are specific emergency numbers and services, such as 190 for the police and 193 for the fire department. In the United States, on the other hand, an emergency can be dialed by typing the number 911.

Calls can originate from many agents (MUKHOPADHYAY et al., 2020). People in need of assistance or other individuals who may have witnessed an incident can report it to the authorities. It is also possible for the first responders to extract incidents from social media or videos. Figure 9 presents the steps that an emergency call goes through until the resource is allocated.

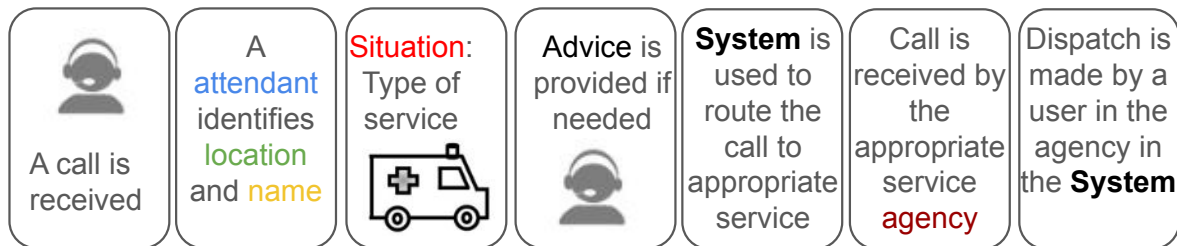


Figure 9 – Pipeline from call to resource allocation for the emergency. Adapted from (MUKHOPADHYAY et al., 2020)

Upon receiving a call, the trained operator must evaluate the situation to determine the necessary response, such as whether it involves a medical emergency or a fire. In instances like airway blockages or CPR, phone instructions are provided before first responders arrive. In Brazil, the operator assesses the needed resources and sends out an emergency team, using a computer-aided dispatch (CAD) system to deploy responders, aided by automatic name and location information (ANI/ALI).

These events define an Emergency Response System and it governs the event response pipeline, including detecting and reporting incidents, monitoring and controlling a fleet of response vehicles, and finally dispatching response teams when the incident occurs (MUKHOPADHYAY et al., 2020). In various cases, there are many government organizations involved, in Brazil, for example, we have the police, firefighters, highway patrol, and traffic managers, among others.

Thus, we can summarize that an emergency management (EM) system is an integrated decision support system composed of various tasks spanning an emergency event's lifecycle. It can be seen that effective prevention of emergency events plays a critical role in the EM system aimed at minimizing or eliminating losses before disaster strikes (HUANG, D. et al., 2021).

Figure 10 illustrates the map interface of the CBMSC dispatcher system (CAD). This interface is utilized throughout the day by numerous users for managing the complete pipeline and dispatching all CBMSC-related incidents, including fires, car accidents, and medical emergencies.

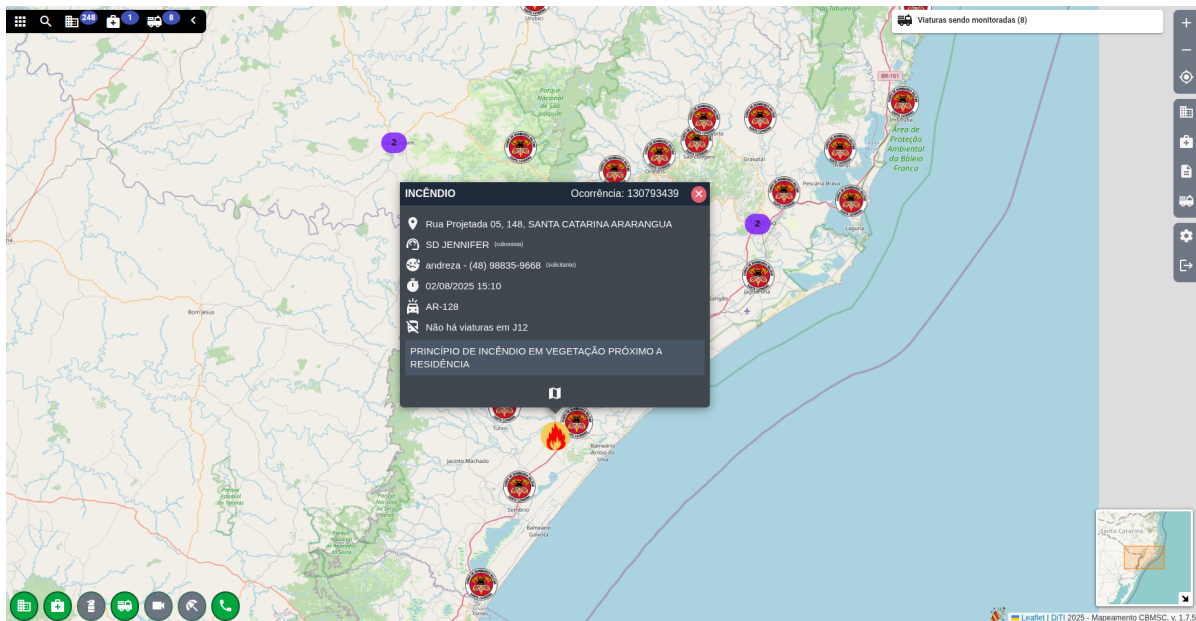


Figure 10 – Map view of the CBMSC dispatcher system interface.

2.3 LIMITATIONS OF POPULATION-BASED STAFFING METRICS IN EMERGENCY SERVICES

Although ratios such as “agents per inhabitant”, whether referring to police officers, firefighters, or paramedics, are frequently cited in public debates and used for baseline comparisons, the technical and academic literature is largely skeptical of their validity (MISHRA et al., 2020). Researchers consistently argue that these population-based metrics fail to capture the operational realities of emergency services and should be replaced with more sophisticated models grounded in actual workload and demand.

The metric of “police officers per 100,000 residents” is widely used but often misunderstood. While various reference values appear in institutional reports, for example, in Brazil the proportion may range from approximately 1 officer per 472 residents nationally to as high as 1 per 180 in the Federal District, these numbers serve primarily descriptive rather than normative purposes. International comparisons, such as U.S. averages fluctuating between 2.17 and 2.51 officers per 1,000 residents, further illustrate the variability rather than establish standards (SECRETARIAT, 2010).

A central critique in the Brazilian literature concerns the so-called “UN standard”, often cited as 340 officers per 100,000 inhabitants (1 per 294). The United Nations Information Center has formally stated that no such staffing parameter exists,

debunking its use as a recommended benchmark. Studies further demonstrate that per-capita ratios overlook crucial contextual factors: for instance, North Carolina data show stark differences between rural police departments (3.87 officers per 1,000) and rural sheriff's offices (1.81 per 1,000), reflecting differing jurisdictional roles rather than population needs.

Because per-inhabitant metrics ignore variations in demand, geography, crime patterns, and organizational practices, the literature concludes that these ratios are fundamentally inadequate for management purposes. Instead, workload analysis, which accounts for call volume, time spent per incident, proactive patrol requirements, and service expectations, is consistently presented as the most reliable and rational method for determining necessary staffing levels.

For fire departments, the limitations of per-capita measures are even more pronounced. The primary drivers of fire service demand are risk factors (e.g., building types, industrial hazards) and geographical constraints, not the size of the population. While some references cite a National Fire Protection Association (NFPA) (NFPA, 2020) ratio of 1.0 firefighter" per population or building, the academic literature is clear: the NFPA itself warns that these figures do not constitute recommended staffing levels or standards. Instead, they reflect medians drawn from heterogeneous fire departments serving vastly different community risk profiles (BRUSHLINSKY et al., 2017).

International standards recommend provision of 1 ambulance for every 50,000 people to fulfil demand for transporting patients to definitive care facilities in Low and Middle Income Countries (LMICs)((DEBAS et al., 2015; MISHRA et al., 2020). Literature (LI, Y. et al., 2023) acknowledges that certain planning documents still provide indicative values, particularly in the context of EMS. These numbers are not strict standards but rather reference points used by local governments or health agencies for infrastructure planning. For example, EMS resources in Tianjin have expanded significantly in recent years, and planning documents cite approximate ratios: in 2021, the number of medical emergency stations was about 1 per 60,000 population, aligned with a local criterion of 1 per 80,000 population and exceeding the threshold of 1 per 140,000 used in Shanghai (GUI et al., 2012). Similarly, Tianjin had approximately 1 ambulance per 30,000 inhabitants, a ratio consistent with national EMS improvement guidelines from 2020. Comparable figures are reported in major Chinese cities, about 1 ambulance per 30,000 people in Beijing and 1 per 40,000 in Shanghai. International data also reveal wide variability: Turkey reports approximately 1 ambulance per 30,000 residents, while Australia has significantly higher coverage,

with 1 per 4,350 people (HÄSKE; SCHMÖCKEL, 2022; GARCIÉA-OLIVARES, 2015).

Beyond highlighting differences in overall demand, these spatial disparities reinforce the need for specialized and region-sensitive decision-making in the allocation and planning of EMS resources. Yue Li et al. (2023) demonstrates that urban, suburban, and rural areas exhibit not only distinct volumes of EMS utilization but also fundamentally different demographic profiles, incident types, and growth trajectories. Urban centers, for instance, sustain the highest and most stable demand, driven in part by older populations and a higher prevalence of critical conditions. Suburban and rural areas, however, demonstrate faster rates of EMS demand growth, signaling emerging pressures that could overwhelm existing infrastructures if resource allocation relies solely on static or population-based metrics. Furthermore, disease profiles vary across regions, with some areas experiencing higher proportions of non-specific diagnoses or trauma-related cases, while others face increasing burdens of cerebrovascular and cardiovascular events.

2.4 MAIN CONCEPTS

This section consolidates the central concepts that support the methodological choices made throughout this thesis. Time series prediction constitutes the core analytical task of this work. Among the possible neural architectures, we adopt LSTM and TFT models due to their established performance and alignment with the characteristics of emergency demand forecasting. LSTM networks remain a strong baseline for nonlinear temporal modeling, particularly when long-term dependencies and irregular patterns are present. TFT, in turn, offers two crucial advantages for this thesis: it natively supports multi-step forecasting and seamlessly integrates multimodal data (static, past-known, and future-known covariates), which aligns directly with the structure of the datasets used in our experiments. These capabilities make TFT especially suitable for the heterogeneous spatio-temporal nature of emergency events.

Spatial regionalization also plays a fundamental role. Given the geographic dispersion of occurrences and the operational heterogeneity across locations, a segmentation stage is necessary to create meaningful spatial units. Partition-based clustering, particularly K-means, was selected due to its simplicity, robustness, and strong empirical performance in geospatial applications. Its ability to generate compact regions and to scale efficiently to large datasets makes it well aligned with the constraints of emergency data partitioning.

Together, these concepts establish the analytical foundation of the proposed

framework: (1) regionalization to uncover spatially coherent patterns, (2) multimodal temporal learning to integrate diverse sources of information, and (3) multi-step forecasting to anticipate short- and long-term emergency demands. The combination of these components, supported by literature and validated experimentally in this thesis, enables a prediction pipeline capable of both accuracy and interpretability across diverse operational contexts.

3 STATE OF THE ART

This chapter presents an overview of the state of the art, where we find approaches to the emergency event prediction problem. We also present a more specific analysis of the main related works. Additionally, we present some works related to the emergency event prediction theme differentiated by the type of emergency. Finally, we summarize the state-of-the-art results to identify research gaps.

An exploratory search and selection methodology were applied in this study. Several academic search engines, such as Google Scholar, have been used. Keywords “emergency events prediction”, “emergency events” and “forecasting model” are searched in the title, abstract, and full text of journal articles and conference proceedings published in English between 2000 and 2024.

3.1 EMERGENCY EVENTS

According to Di Huang et al. (2021), due to the wide variety of emergency events, including both natural and human-caused ones, the prediction model is specific to each event, considering different social and environmental factors such as socioeconomic conditions and geographic characteristics. Decision-making during disaster management operations also varies considerably depending on the type of event, considering factors such as severity, affected area, population density, and surrounding landscape, among others. Although few, some efforts have been dedicated to discussing current methods of prediction and evaluation of natural disasters, such as the study by (AMEZQUITA-SANCHEZ et al., 2017).

On the other hand, human-caused disasters have also become increasingly important worldwide, such as severe accidents in road and urban transportation systems, and public health emergencies worldwide, among other situations. Therefore, in this work, we aim to fill this gap by reviewing works on emergency event prediction in response to different types of emergency events, including both natural disasters and human-caused ones.

3.1.1 Urban Emergencies

Emergencies occurring in urban areas are often caused by two types of events: natural disasters and social incidents (e.g., fire, explosion, terrorist attack, crime, and traffic accident). The affected people require immediate assistance and must be evacuated from the risk area. Therefore, the main objective of the urban emergency demand

prediction problem is to determine the number of victims or occurrences, providing information for evacuation planning and optimization of resource management, including shelter location and capacity, evacuation route, and resource allocation, among others (HUANG, D. et al., 2021).

3.1.1.1 Crime

Predicting the temporal and spatial distributions of crimes in urban areas is an old issue in urban management systems (ZHAO; TANG, 2017). In the literature, it is generally assumed that the number of crimes depends on explanatory variables such as unemployment and income, based on which linear regression models are conducted to analyze their interrelationships (ALVES et al., 2015).

However, Alves et al. (2018) indicates that the relationship between crime occurrence and urban indicators is often nonlinear. This study then uses the Random Forest algorithm to quantify the importance of indicators and predict the number of homicides, achieving an accuracy of 97%. Chao Zhang et al. (2015) point out that most crimes are opportunistic (not necessarily planned by the actors), but the location and timing of crimes depend on policing. Therefore, this study develops a patrol scheduler to predict crime occurrence by learning criminal behavior, analyzed through a multi-modal database.

Seo et al. (2018) propose an automatic crime classification system. A partially generative neural network is introduced, capable of classifying crimes with complete or partial information. Gholami et al. (2018) focus on the method of predicting crimes that occurred in the wildlife protection domain. To overcome the lack of observed data, a new ensemble technique is designed to predict poachers' behavior, whose results are subsequently used in the patrol planning problem.

3.1.1.2 Traffic Accidents

Predicting traffic accidents is another important use of prediction models in urban emergencies. Numerous research works apply statistical and artificial intelligence approaches to analyze the factors influencing traffic accidents and to predict their occurrence. With the help of data collection techniques, a large amount of data from various sources is available for traffic flow analysis. To select the most significant variables, (LIN et al., 2015) propose a new variable selection method based on the Frequent Pattern (FP) tree. The result shows that the proposed FP method is better than the Random Forest-based model in terms of prediction model type.

(LU et al., 2015) focus on predicting accident hotspots in urban areas. A logistic regression analysis investigates the relationships between traffic accidents and road type, vehicle type, driver's condition, etc.

Many efforts have been dedicated to neural network approaches. For example, Alkheder et al. (2017) uses an artificial neural network (ANN) to predict the severity of traffic accident injuries based on nearly 6000 traffic accident records. Considering the case with limited data, (YUAN et al., 2018) developed a deep learning approach for traffic accident prediction, namely the Convolutional Long Short-Term Memory (ConvLSTM) neural network model.

3.2 OVERVIEW OF THE STATE OF THE ART

Various methodologies have been proposed for predicting emergency-related incidents over the years (MUKHOPADHYAY et al., 2020). In one instance, a spatial approach using ANN optimized with an evolutionary algorithm and observations related to occurrences in Athens, Greece was utilized to predict incident locations (GREKOUSIS; LIU, 2019). This method enabled the efficient allocation of resources, thereby reducing response time. In (SHARMA et al., 2024) the authors use a data-driven approach by analyzing the associations of event types with socioeconomic and demographic data at the neighborhood level, identify a set of predictors for each event type, and use a Negative binomial as a regressor.

The study outlined in (CORTEZ et al., 2018) focuses on predicting emergency events using spatial and temporal features. The authors highlight that their research marks the first instance of employing LSTM for emergency event prediction. Their analysis was conducted on data sourced from the Ministry of the Interior website of Guatemala, provided by the national police. In their investigation, the authors utilized the LSTM model to contrast with traditional time series techniques, such as ARIMA, MA, and WMA. Additionally, they employed conventional Machine Learning methods, including DT, SVM, and MLP. The authors elaborate on the parameter tuning process undertaken to ensure a fair comparison with LSTM, adjusting the parameters of this model accordingly. Furthermore, the paper introduces in the context of emergency events both binary prediction (occurrence or absence of emergency events per day) and regression (number of daily occurrences) approaches. To evaluate the predictive capacity of the model, the authors prepared the dataset in two different manners, assessing the independence or dependence of regions on the model's predictions.

Additionally, (MUKHOPADHYAY et al., 2019) proposed an online approach

to predict emergency incidents using an RNN, coupled with a Decision Theoretical Framework that included prediction, environment simulation, and dispatch. This approach could consume streaming data and efficiently update existing incident arrival models. In (HADAD et al., 2019), a model to predict firefighter interventions in France was proposed, considering the selection of relevant variables. The input data consisted of 44000 items, with 58 features, including month, day of the month, and weather conditions, among others.

Unsupervised learning methods can also be employed to understand patterns in data that are not labeled. (MUKHOPADHYAY; VOROBAYCHIK, 2017) used a hierarchical clustering approach with data from traffic accidents in Nashville, United States, and concluded that the results could reduce incident waiting times, providing a structured and systematic way of dispatching. Furthermore, (PETTET et al., 2017) used Similarity Based Agglomerative Clustering (SBAC) to categorize incidents and Bayesian Networks as an inference technique, using data from the Nashville Fire Department. This combination demonstrated the potential to predict incidents occurring in a large geographical area accurately.

Social networks can also be used to extract valuable information regarding incidents. In (ZHANG, Z. et al., 2018), 3 million tweets related to traffic accidents in the Virginia region and New York City over one year were evaluated using Deep Belief Network (DBN) and LSTM to extract tokens. The authors suggested that social networks could be used in conjunction with traditional methods for accident detection. (YE et al., 2022) uses a set of traditional Machine Learning methods to evaluate the prediction of structural displacement.

3.3 MAIN RELATED WORK

This section presents the main related works that were investigated to this work.

3.3.1 ***An architecture for emergency event prediction using LSTM recurrent neural networks***

The work presented in (CORTEZ et al., 2018) addresses the problem of predicting emergency events using spatial features, such as the locations of emergency events, as well as temporal characteristics. The authors emphasize that this study was the first to use an LSTM in emergencies. The experiment was applied to data obtained from the Ministry of the Interior website of Guatemala, published by the national police.

The authors used the LSTM model to compare traditional time series techniques, such as ARIMA, Moving Average (MA), Weighted Moving Average (WMA). In addition, traditional Machine Learning techniques were used, including Decision Tree (DT), SVM, and MLP. The authors describe how they tuned the parameters of the algorithms to make the comparison with LSTM fair, as adjustments were made to the parameters of this model.

Figure 11 presents the workflow that the authors employed in the data. From raw data in a database containing event location and event count, preprocessing involved using a clustering technique and temporal aggregation of events by locality, in addition to windowing. These data are then passed to the LSTM model, which returns as output the future prediction based on the desired windowing.

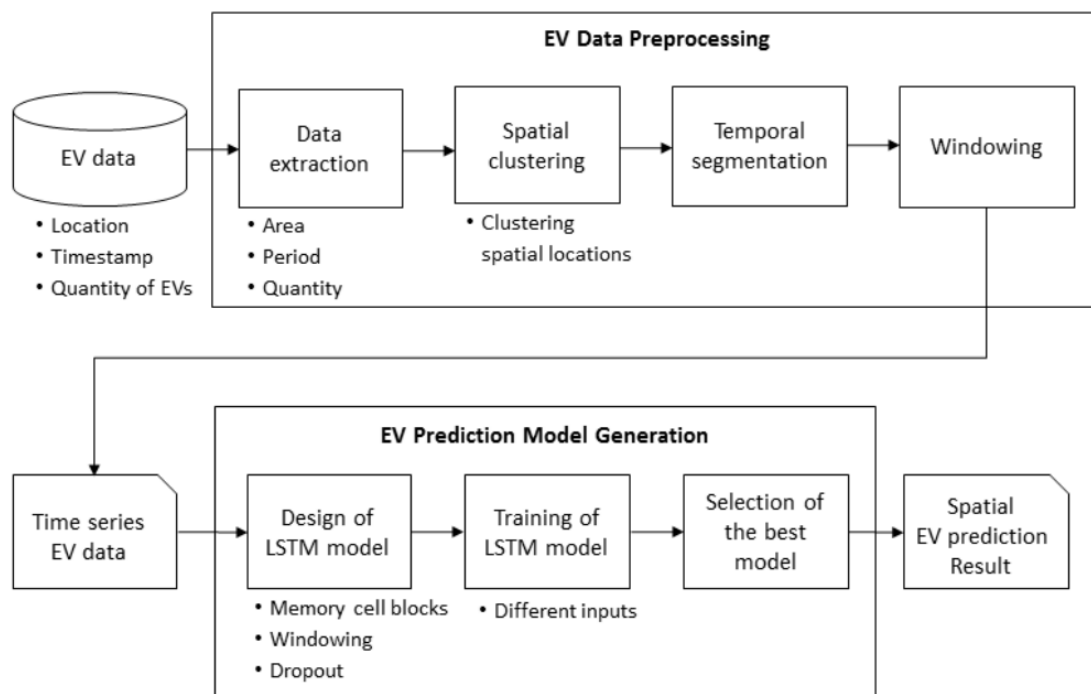


Figure 11 – Experiment blueprint. Source: (CORTEZ et al., 2018)

Another point to highlight in the work is the approach to the problem. First, after clustering the regions in space, the authors divided the data into four types. A binary approach and a regression approach, in addition to considering the dependence (considers all clusters in prediction and training) and independence (considers only one cluster at a time in prediction and training) between the regions in space. The two modeling approaches, regression/binary, can be visualized in Figure 12.

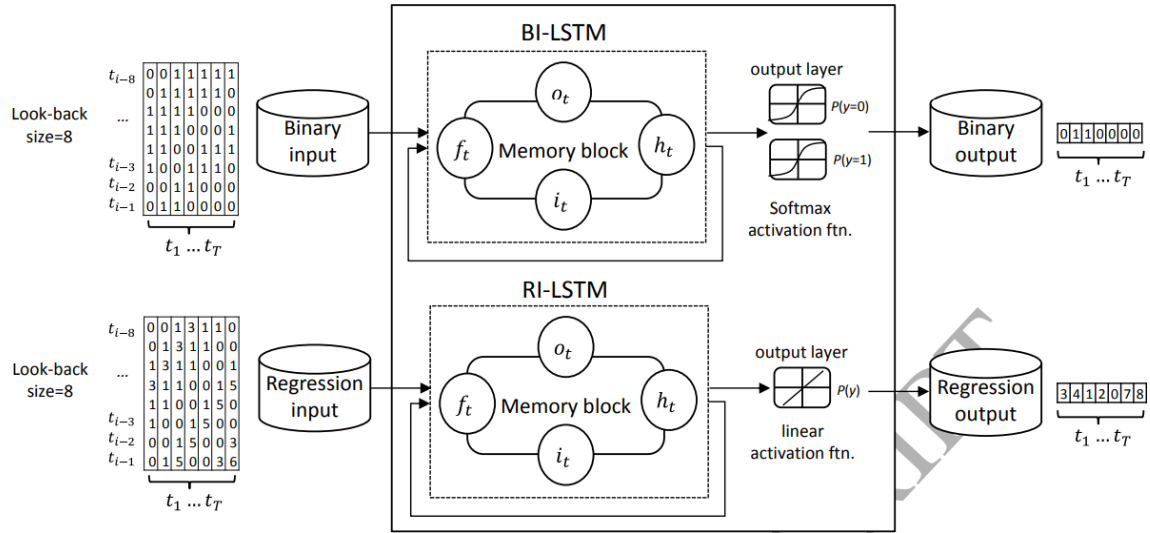


Figure 12 – Two LSTM Models considering independence between clusters. Source: (CORTEZ et al., 2018)

3.4 SUMMARY AND RESEARCH OPPORTUNITIES

In this section, the works summarized in a table covering different theoretical and methodological approaches will be presented. The main findings and contributions of the listed works will be discussed. This analysis will enable a better understanding of the limitations and possibilities for expanding the topic, contributing to formulating new research questions.

In Table 1, the most relevant works for the research are summarized. These works are differentiated by the type of data used to predict emergency events, the outcome used to predict, the methodology proposed for obtaining the results, the multimodality (MM), if the work uses Multistep (MS) to predict more than one step ahead, and the use of spatial segmentation (SS).

As mentioned earlier, the work in (GREKOUSIS; LIU, 2019) presents an approach to emergency event localization and resource allocation. The approach differs from those used in other works that present specific time series techniques. In (CORTEZ et al., 2018), the authors present a space-time approach of LSTM compared to standard time series prediction methods. A related approach is seen in (VIEIRA ROQUE et al., 2025), where the authors introduce a novel multistep prediction framework. Their study compares region-specific models to generalized ones, showing that tailored, region-specific models tend to outperform generalized approaches in terms of prediction error.

In (MUKHOPADHYAY et al., 2019), the authors present another approach to prediction and resource allocation for emergency events. The contribution of this work is found in a framework that comprises incident prediction, environment simulation, and resource allocation. On the other hand, the work in (HADAD et al., 2019) presents an approach to identify the importance of attributes in event prediction. (PAYARES-GARCIA et al., 2023) proposes a dynamic spatio-temporal stochastic modeling approach to identify complex underlying processes of crime data coming from calls to the emergency phone in an urban environment over 10 years. Authors use doubly stochastic Poisson processes to model data (REINHART, 2018) and evaluated this approach in making predictions over the model.

Another important work in this review is found in (MUKHOPADHYAY; VOROBEYCHIK, 2017), where the authors present a clustering approach to maximize the coverage of responses by finding the optimal location for dispatch. The work of (PETTET et al., 2017) also presents a resource allocation approach using a clustering algorithm based on past accidents and the number of road intersections. In (ZHANG, Z. et al., 2018), we find an approach of social media data (Twitter) analysis and Deep Belief Network as predictors of events.

In a literature review, the work (HUANG, D. et al., 2021) addresses different methods in disaster prediction methods for emergency occurrence and resource demand. The article aims to provide a comprehensive survey of prediction technologies widely used in emergency management (EM) and discuss the characteristics of each type of emergency event. The work distinguishes events into different categories, caused by humans and natural disasters. Among the findings of this work, it is indicated that building a real-time monitoring and prediction system is still an open issue in the literature. Additionally, the authors also point out that prediction methods integrating prediction and resource allocation need more attention.

Work	Data	Outcome	Methodology	MM	MS	SS
(GREKOUSIS; LIU, 2019)	2851 emergency events	Location of the incident	Evolutionary algorithm to optimize the Neural Network.	-	-	✓
(CORTEZ et al., 2018)	Occurrences of 1121 days	Number of occurrences	LSTM model compared with ARIMA, MA	-	-	✓
(MUKHOPADHYAY et al., 2018)	9345 incidents	Number of occurrences	Online architecture for prediction emergency and dispatch of teams	-	-	-
(HADAD et al., 2019)	58 variables (day of the year, holiday, weather conditions, etc.)	Number of occurrences	Linear regression to evaluate the effectiveness of each variable	✓	-	-
(MUKHOPADHYAY; VOROBAYCHIK, 2017)	20148 traffic Accidents	Number of occurrences	Clustering algorithm based on past accidents and the number of road intersections evaluated	-	-	✓
(PETTET et al., 2017)	Two years of incident with 477837 incidents	Cluster location	Similarity Based Agglomerative Clustering (SBAC) to categorize incidents and Bayesian Networks for inference	-	-	-
(ZHANG, Z. et al., 2018)	Related Tweets with accidents	Prediction of accident	Deep Belief Network (DBN) as a predictor	✓	-	-
(SHARMA et al., 2024)	5 years of occurrences	Number of Occurrences	Extensive data analyses. Negative binomial regression model	-	-	✓
(YE et al., 2022)	Temperature	Prediction of structural displacement	Different traditional ML approaches	-	-	-
(PAYARES-GARCIA et al., 2023)	10 years of occurrences	Number of Occurrences	Doubly stochastic poisson processes to model data	-	-	✓
(VIEIRA ROQUE et al., 2025)	5.5 years of occurrences	Number of Occurrences	LSTM multi Spatio-Temporal approach	-	✓	✓
This work	Emergency events data	Number of Occurrences	General methodology for time series prediction emergency events	✓	✓	✓

Table 1 – Summary table with the main works and proposed solution (MM = multimode, MS = multistep, SS = spatial segmentation).

In this section, we presented distinct approaches from literature to predict emergency events that share some similarities. All works aim to predict emergency events to assist in planning (for all works) and allocation (in some) of emergency resources. All works use machine learning techniques for event prediction and significantly contribute to predicting emergency events. The studies use different data sources for event prediction, including social media data (Twitter), data from past accidents, sensor data, and simulation data. Additionally, the works employ different theoretical and methodological approaches to event prediction, including time series-based approaches, clustering approaches, and LSTM space-time approaches.

The works also have different outcomes and objectives, such as maximizing response coverage, identifying attributes' importance in event prediction, and allocating resources for emergency events. These differences demonstrate that there are several possible approaches to predicting emergency events. From the identification of the approaches of the works, we can identify two major research gaps: resource allocation and event prediction. These two research spaces are discussed in some of the mentioned works. An important contribution to advancing the theme could be a system that uses predictions as input for resource allocation, which is still an open question (HUANG, D. et al., 2021).

While the literature on the subject is often limited to specific regions or single data types, our work advances the field by conducting a comprehensive case study using multiple categories of emergency occurrences, including firefighting and medical incidents, to support more effective resource allocation. The principal contribution of this thesis is the integration of spatial segmentation, multimodal data fusion, and multi-step prediction into a unified and coherent framework. By combining these three dimensions, our approach addresses several open challenges in emergency forecasting, enabling the construction of region-aware, context-enriched, and horizon-flexible predictive models. This integrated framework not only improves upon existing methods but also offers a structured foundation that can serve as a reference model for future researchers and practitioners seeking to implement data-driven emergency event prediction systems. The next chapter (Chapter 4) presents the proposed framework in detail, outlining each of its components and illustrating how they operate cohesively to enhance prediction accuracy and decision-making.

4 PROPOSED SOLUTION

This chapter presents an overview of the framework proposed in this work, called FORESEE. The tools and methods are described in an interconnected manner to generate the expected results. Our prior analysis (see Chapter 3) highlighted several key limitations and research gaps in existing emergency event prediction approaches:

- The inherent difficulty in predicting emergency events both locally and temporally, largely due to the unique characteristics of different spatial areas. This underscores the need for region-specific models rather than generalized ones, which tend to lose localized specificities;
- The absence of automated processes for defining and constructing these region-specific models, a task that can be time-consuming and susceptible to errors if performed manually;
- The challenge of integrating diverse data sources, as single acquisition methods often fail to provide an understanding of complex natural processes and environments;
- A prevalent limitation in the literature regarding multi-step-ahead forecasting, which is crucial for short, medium, and long-term planning in emergency management and resource allocation.

To systematically tackle these identified challenges, the FORESEE framework proposes a data-centric approach. Automates critical steps within the time-series modeling process, providing mechanisms for:

- Regionalized modeling using spatially-segmented data: The framework segments original time series into specialized datasets based on geographical coordinates (latitude and longitude);
- Handling of multimodal data types: FORESEE incorporates external data sources such as weather conditions, traffic information, and public event schedules, enriching the primary emergency event data. This process can uncover subtle patterns and correlations that might be missed when data sources are analyzed in isolation.

- Multi-step forecasting: The framework supports the prediction of emergency events across various future time windows, providing insights for strategic resource allocation and long-term preparedness.

4.1 THE FORESEE FRAMEWORK

The *FORESEE* framework is designed to enable accurate and scalable emergency event prediction by integrating spatial segmentation, multimodal data, and advanced temporal modeling. It is organized into five key stages (illustrated in Fig. 13), each contributing distinct capabilities to support predictive performance, regional adaptability, and model interpretability.

The phases are structured as follows. **A) Data Collection:** The process begins by defining macro-regions and visualizing emergency occurrences using shapefiles. These regions, characterized by clusters of cities with shared operational and geographic characteristics, are analyzed through spatial distribution maps of event categories. Preliminary exploratory clustering is applied to uncover heterogeneous emergency patterns across the territory. **B) Regionalization:** Next, the framework performs fine-grained spatial segmentation by applying clustering algorithms (e.g., K-Means, DBSCAN, Agglomerative) to group occurrences into smaller, behaviorally homogeneous zones. This regionalization step aims to create region-specific models that capture localized emergency dynamics, improving model precision and generalization. **C) Multimodal Data and Spatio-Temporal Modeling:** In this phase, multiple heterogeneous data sources are used, including meteorological, calendrical, and operational attributes, forming a multimodal (MM) dataset. These features are input to the model, which is capable of handling both static and dynamic covariates while enabling multi-step (MS) time series forecasting. **D) Time Series Analysis and Preparation:** Before training, time series analysis is conducted to explore structural patterns. This includes decomposing series into trend, seasonality, and residual components, testing for stationarity using the Augmented Dickey-Fuller (ADF) test, and inspecting autocorrelation via ACF and PACF plots to guide model design. **E) Evaluation and Interpretability:** Finally, the trained model is evaluated using metrics such as MAE and RMSE across different forecasting horizons. The framework offers interpretability by highlighting influential features over time. Comparative experiments are also conducted with traditional forecasting models (e.g., ARIMA, STL), classical machine learning methods (e.g., Decision Trees, SVM), and LSTM-based architectures.

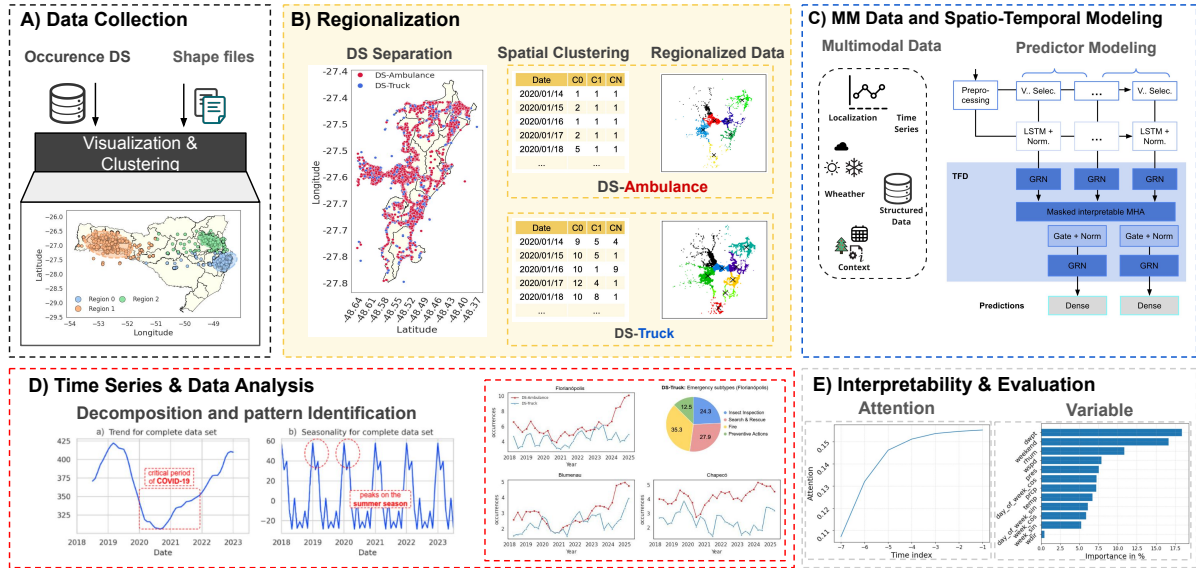


Figure 13 – Proposed framework.

The primary objective of this framework is to produce accurate emergency event forecasts through an integrated time-series modeling pipeline. It addresses key challenges in emergency prediction by combining multimodal data aggregation, spatial segmentation, preprocessing, and predictive modeling. This hybrid approach captures short- and long-term temporal patterns while incorporating relevant contextual information. Within this framework, the main contributions lie in steps B, C, and E: (B) confirming that regionalized models outperform generalized ones; (C) demonstrating that multimodal data improves predictive performance; and (E) enabling multi-step forecasting with interpretable outputs. Together, these elements offer a structured and effective foundation for improving preparedness, resource allocation, and strategic decision-making in emergency response.

To validate this framework, we conducted experiments in two stages. In the first set of experiments (see Chapter 5, Section 5.1.1), we evaluated components A, B, and part of C, focusing primarily on spatial segmentation and demonstrating that regionalized models outperform generalized ones. In the second stage (see Chapter 5, Section 5.2.1), we implemented the complete pipeline, components A, B, C, D, and E, to achieve the full capabilities of the framework, including multimodal data integration and multimodal data. Although the complete workflow confirms the effectiveness of the proposed design, each component can also be applied independently, allowing researchers and practitioners to adapt the framework to different datasets, operational constraints, or specific modeling needs.

4.1.1 Data Collection

The data collection phase begins with defining the spatial context in which the predictive models will operate. This involves selecting a specific area, such as a city, district, or region, to serve as the basis for model training and evaluation. The choice of spatial scope can be guided by domain experts with experience in emergency response patterns or based on administrative boundaries such as city limits. Our experimental results demonstrated that both macro segmentation and predefined city-based segmentation are effective for generating reliable predictive models (VIEIRA ROQUE et al., 2025).

After the spatial scope is determined, either by experts or existing city limits, relevant data are gathered and structured. These include records of emergencies, weather conditions, calendar events such as holidays, and other contextual elements. A fundamental requirement for emergency records is that they be time-stamped and include at least some form of spatial reference, as presented in Table 2. Geolocation is essential for associating events with specific regions and enabling effective spatial segmentation. In cases where precise coordinates are missing, alternative location attributes such as neighborhoods or street addresses can be used to infer the event's location. This preprocessing step ensures that spatial gaps do not compromise model quality.

Table 2 – Real occurrence data.

ID	Date	Longitude	Latitude
2020001	01/01/2020	-48497971	-27546850
2020233	21/01/2020	-48497967	-27546857
2020456	11/01/2020	-48497956	-27546789
2020249	25/01/2020	-48497890	-27546798
2020598	29/01/2020	-48497789	-27546878
2020489	12/01/2020	-48497876	-27546760

In addition to defining the spatial and temporal aspects of the data, it is crucial to ensure diversity in the types of emergency occurrences collected. Different emergency categories, such as traffic accidents, medical assistance, and fire-related incidents, may follow distinct temporal patterns, be influenced by other external factors, and require different response strategies. Therefore, separating and modeling these categories individually allows for the development of more specialized and accurate predictors. Predictive modeling for emergency events depends on the availability of a sufficiently long historical dataset. Evidence gathered from our case studies suggests

that a historical window of at least five years offers sufficient variability and stability to capture seasonal effects, long-term trends, and rare occurrences.

Finally, data cleaning is an essential part of this step. Because emergency management systems may malfunction, the raw data often contain duplicated entries referring to the same incident, as well as outliers such as an unusually high number of events recorded on a single day, values that exceed what a region is realistically capable of attending. Our experimental analysis confirmed the presence of these anomalies. Therefore, the data must be cleaned to remove duplicates, correct inconsistencies, and filter implausible outliers to ensure a reliable foundation for subsequent modeling. Additionally, at this stage the user must define the prediction granularity, daily, weekly, hourly, or any other temporal window. This choice depends strongly on the availability and quality of the data: finer granularities (such as hourly prediction) require consistent, high-resolution timestamps, while coarser ones (such as daily or weekly) are feasible even with lower temporal precision. Ensuring this alignment between data quality and temporal granularity is essential for building reliable forecasting models.

4.1.2 Regionalization

The regionalization process begins with the spatially-tagged emergency data obtained in the Data Collection phase. Since the goal is to train region-specific predictive models, it is essential to segment the geographical area into meaningful sub-regions. This segmentation is performed based on the geographic coordinates (longitude and latitude) of the recorded occurrences, which are grouped using clustering algorithms to form region-specific clusters denoted as $c_0, c_1, \dots, c_n$. Each cluster corresponds to a localized region expected to exhibit unique behavioral patterns in emergency events.

To ensure that the segmentation reflects meaningful regional structures, we evaluate the quality and coherence of the clusters using metrics such as the Silhouette score (ROUSSEUW, 1987). The optimal number of clusters can also be informed by domain experts, for example, by considering the number of local emergency response organizations or operational bases in a given area.

Figure 14 illustrates the applied methodology for spatial segmentation. The clustering approach aims to capture the intrinsic characteristics of each region, allowing the predictive models to adapt to localized dynamics and improve forecast accuracy.

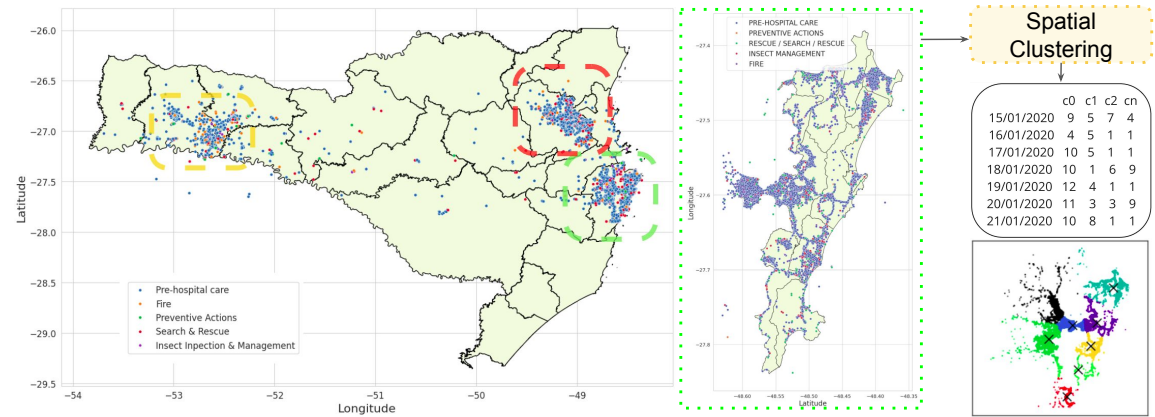


Figure 14 – Methodology to segment the regions.

The output of the regionalization process is a set of n distinct time series, each associated with one of the spatial clusters $c_0, c_1, \dots, c_n$ identified through the segmentation process. These time series are constructed by aggregating the number of emergency occurrences within each cluster over fixed time intervals (e.g., daily), resulting in time-stamped clusters aggregated by quantity of occurrences. This aggregation step transforms spatial clusters into temporal sequences, enabling the modeling of how emergency events evolve over time in specific regions.

These region-specific time series are specialized representations that capture the temporal evolution of emergency events within spatially coherent areas. By segmenting the data spatially and then modeling each segment as a separate time series, the framework effectively integrates both spatial and temporal dimensions of the phenomenon. This enables the construction of localized predictors that are sensitive to the unique behavior of emergencies in each region, ultimately improving the precision and interpretability of the forecasting models.

Table 3 presents an example of the output produced during the regionalization stage of the framework. After applying the clustering algorithm, each emergency event is assigned to a spatial cluster, forming the basis for region-specific time series. The table shows the cluster label and the corresponding number of occurrences for each date, illustrating how raw event data is transformed into spatially segmented series.

Table 3 – Data result with cluster assignments and occurrence counts.

Date	Cluster Label	Occurrences
2018-01-01	0	10
2018-01-01	1	10
2018-01-01	2	6
2018-01-01	3	9
2018-01-02	0	12

4.1.3 Multimodal Data and Spatial-Temporal Modeling

Handling data is a critical step in enhancing the predictive capacity of the framework by incorporating heterogeneous and contextually relevant information. The motivation for integrating multimodal data lies in obtaining a comprehensive and unified view of the system, enabling better decision-making, exploratory analyses, and modeling of complex relationships between external factors and emergency events (LAHAT et al., 2015).

At this stage, multiple external data sources are collected, processed, and integrated with the region-specific time series produced in the regionalization phase. These sources include, but are not limited to, weather conditions, traffic indicators, special calendar days (e.g., holidays and public events), and infrastructure-related factors. A key requirement for all supplementary data is that it must be time-stamped, ensuring temporal alignment with the emergency occurrences.

Due to the regionalized nature of our dataset, each external data source must also be *spatially aligned* to the corresponding clusters $c_0, c_1, \dots, c_n$, where n indexes spatial positions. For example, weather and traffic data must reflect conditions specific to the geographic region defined by each cluster. This ensures that the model learns meaningful localized patterns, rather than diluted or generalized signals. The fusion process, therefore, uses the dates and spatial assignments of the clusters as anchors to merge datasets on a per-region, per-timestamp basis.

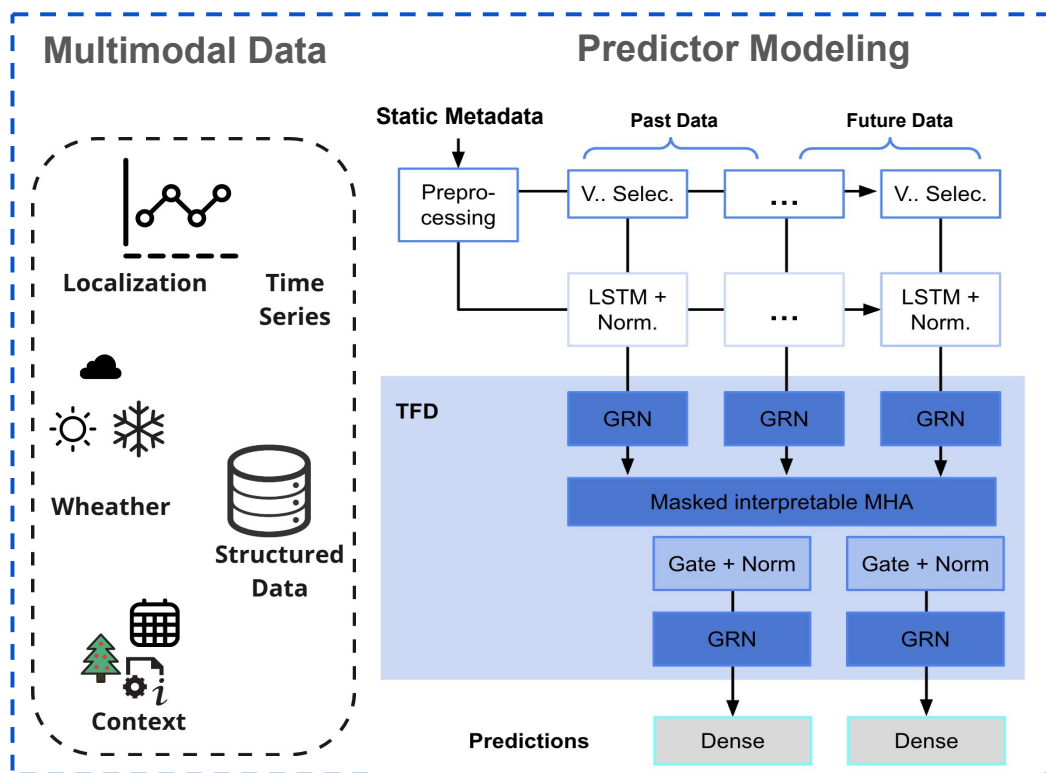


Figure 15 – Multimodal data and predictor modeling approach.

Figure 15 illustrates the complete process of multimodal data and spatial-temporal modeling. After aligning the datasets, we proceed to structure the input features to feed a predictive model capable of handling such heterogeneity. In the figure, we highlight the use of the TFT, a model that is particularly well-suited to this task due to its native ability to process static, known, and observed variables, while leveraging attention mechanisms to focus on the most relevant temporal dependencies.

However, the framework remains flexible and model-agnostic: the TFT can be replaced with other sequential models such as LSTM networks, depending on the context and availability of features. Regardless of the chosen model, the key requirement is that it must be capable of learning from multivariate temporal sequences that integrate both emergency counts and contextual variables.

By incorporating multimodal data in a region-aware and temporally precise manner, this step enables the construction of robust predictors that are sensitive to a wide range of dynamic conditions. The resulting models are thus able to anticipate

changes in emergency event patterns not only based on past occurrences but also based on influencing external factors, enhancing forecast accuracy, responsiveness, and operational value.

Multi-step prediction is important to this framework, as emergency managers require forecasts not just for the next day, but for multiple days ahead to plan staffing, anticipate peak demand, and deploy resources preventively. Different model families handle this requirement differently: autoregressive deep learning models such as LSTMs often require explicit multi-step strategies (e.g., recursive prediction or direct multi-output training), whereas transformer-based models can natively predict several steps ahead within a single forward pass. For this reason, the multimodal modeling stage of the framework incorporates multi-step forecast preparation directly into the feature engineering and model design, ensuring that predictions remain robust across different horizons.

Table 4 illustrates the next enrichment step, in which regionalized data is augmented with weather variables. These weather attributes—such as temperature, humidity, precipitation, and wind conditions, are aligned both temporally and spatially with each cluster. This integration enables the framework to incorporate environmental factors that may influence emergency event patterns.

Table 4 – Multimodal integration with weather variables added to regionalized data.

Date	Cluster	Occ.	Temp	Dwpt	Rhum	Prcp	Snow	Wdir	Wspd	Pres
2018-01-01	0	10	24.7	20.0	75	0.0	–	111	4.3	1010.1
2018-01-01	1	10	24.7	20.0	75	0.0	–	111	4.3	1010.1
2018-01-01	2	6	24.7	20.0	75	0.0	–	111	4.3	1010.1
2018-01-01	3	9	24.7	20.0	75	0.0	–	111	4.3	1010.1
2018-01-02	0	12	26.0	23.1	84	0.0	–	14	6.5	1007.1

Table 5 shows the final multimodal dataset in compact form. In addition to the regionalized information and weather variables, this table includes calendar-based features (e.g., holidays, weekday, month) and temporal indexing variables required for supervised forecasting. Together, these elements form the complete multimodal input structure used by the predictive models in Stage C of the framework.

Table 5 – Complete multimodal table with temporal and calendar features added.

Date	Cluster	Occ.	Weather Vars	Holiday	DOW	Month	Half-Day	t_idx
2018-01-01	0	10	...	Christmas	0	1	AM	00
2018-01-01	1	10	...	Christmas	0	1	AM	00
2018-01-01	2	6	...	Christmas	0	1	AM	00
2018-01-01	3	9	...	Christmas	0	1	AM	00
2018-01-02	0	12	...	No Holiday	0	1	AM	01

4.1.4 Time Series and Data Analysis

This step plays a fundamental role in the framework, as the quality of the modeling process directly depends on how well the temporal structure of the data is understood and processed. Beyond merely preparing the data for forecasting, time series analysis serves to reveal complex patterns, identify anomalies, and offer potential explanations for phenomena observed in model behavior. For instance, a recurring increase in emergency occurrences during the summer in coastal regions may be detected and better explained through careful seasonal analysis.

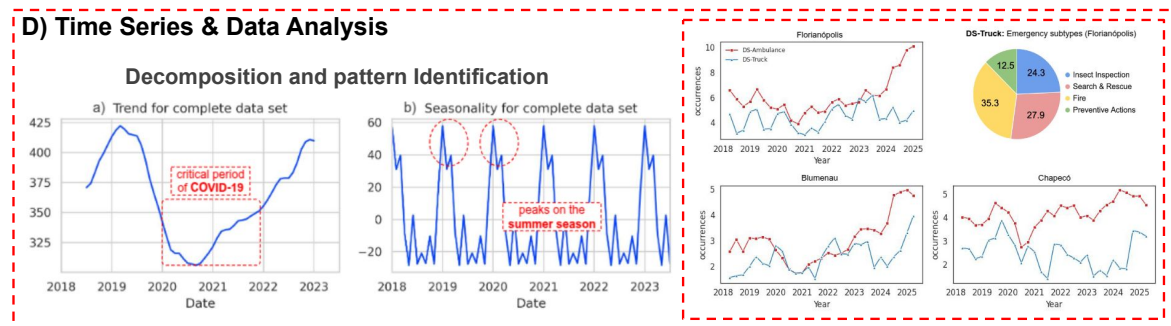


Figure 16 – Time Series and Data Analysis.

As shown in the red dotted section of the framework (Figure 16), we begin this step by analyzing autocorrelation patterns. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are applied to quantify the dependencies between current and past values of the time series. ACF measures the overall correlation at various lag intervals, while PACF isolates the direct effect of specific lags, controlling for intermediate steps. These tools are important for identifying the presence of seasonality, repeating cycles, or delayed dependencies in emergency data. This is particularly relevant for capturing weekly, monthly, or even yearly behavior patterns (VIEIRA ROQUE et al., 2025).

To ensure the validity of these analyses, we first assess whether the series is stationary, a prerequisite for many time series models. For this purpose, we apply the

Augmented Dickey-Fuller (ADF) test (FULLER, 1976), a widely used statistical test that evaluates the null hypothesis that a unit root is present in the data. A strongly negative ADF statistic suggests the series is stationary and suitable for modeling without further transformations.

Additionally, we apply time series decomposition techniques to disentangle the observed series into its fundamental components: trend, seasonality, and residuals. The trend captures long-term directional movements (e.g., a steady increase in emergency calls over the years), while seasonality highlights periodic patterns, such as weekend spikes or seasonal effects related to holidays or tourism. The residual component represents the irregular fluctuations or "noise" that cannot be explained by trend or seasonality alone.

This decomposition is not only useful for modeling but also for conducting descriptive and exploratory data analysis, helping to understand underlying structures or irregular behaviors in emergency occurrences. By interpreting these components separately, the framework is better equipped to:

- Justify modeling decisions (e.g., setting prediction windows),
- Diagnose poor model performance (e.g., unexplained variance in residuals),
- Highlight significant phenomena embedded in the data (e.g., vacation seasonality).

This step provides the analytical evidence required to make informed choices about both data preparation and model design. By quantifying temporal dependencies, seasonality, stationarity, and unexplained variance, the time series analysis stage supports decisions such as the selection of prediction windows, the need for transformations or filtering, and the identification of meaningful behavioral patterns in the data. The result is not the model itself, but a clearer understanding of the data-generating process—knowledge that guides all subsequent modeling and evaluation activities.

4.1.5 Interpretability and Evaluation

The final module of the proposed framework focuses on the evaluation and interpretability of the predictive models. This step is essential not only for both validating model performance and elucidating prediction mechanisms, thereby enhancing transparency and reliability of outcomes, particularly in critical areas like emergency response. Figure 17 presents this step.

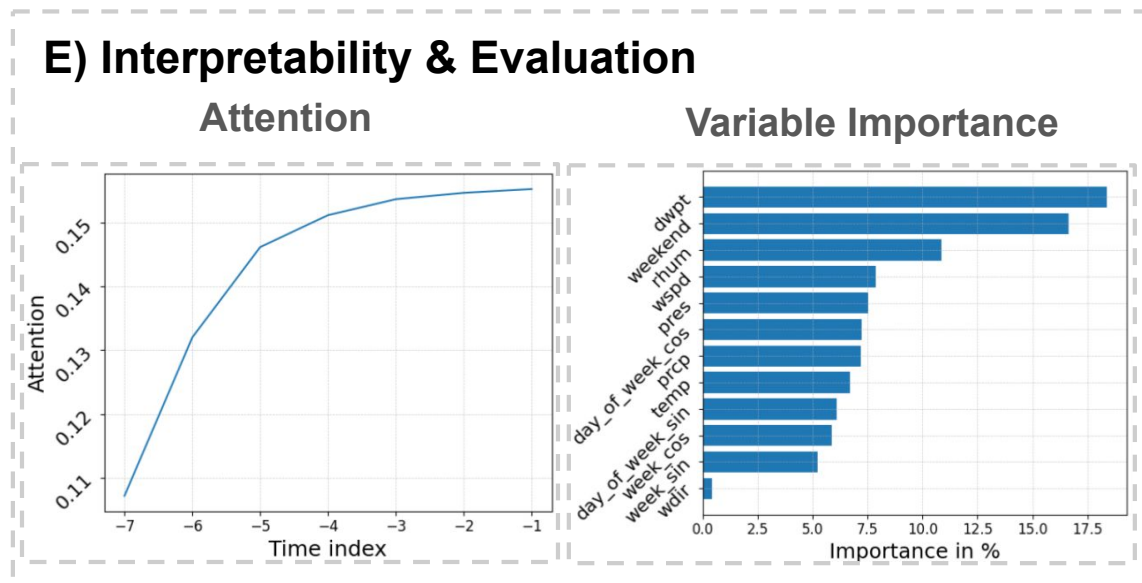


Figure 17 – Evaluation and Interpretability.

The process comes from the construction of the predictive model, a task that often involves managing many hyperparameters. Given the complexity and sensitivity of these parameters to the data distribution, we employ hyperparameter tuning strategies, such as grid search, random search, or more sophisticated optimization techniques, to identify suitable configurations. Emergency event patterns vary substantially across regions, and each spatial cluster may exhibit unique temporal, behavioral, and contextual characteristics. For this reason, the framework is designed to search for the best configuration adapted to the local characteristics of each region, rather than imposing a global parameterization. These locally tuned parameters are carefully tracked and evaluated within the context of the macro-regional segmentation defined in previous steps, ensuring that the resulting models are both specialized and aligned with regional dynamics, thereby maximizing predictive performance.

The predictions must be evaluated across different forecasting horizons. We analyze how model performance evolves with the prediction length to assess the accuracy of both short-term and long-term predictions. This is crucial in our context, as anticipating emergency occurrences multiple steps into the future is vital for strategic planning and resource allocation.

In parallel, we assess model interpretability, an important requirement for real-world deployment. For models like the TFT, attention mechanisms and variable importance scores can be analyzed to identify which features most strongly influence predictions at different time steps. Such interpretability tools allow us to answer criti-

cal questions, such as whether weather, traffic, or temporal markers (e.g., weekends, holidays) drive model outputs, and how these effects differ across regions or time windows.

Overall, this step ensures that the predictive models are not treated as black boxes. By combining evaluation metrics with interpretability analyses, we support both technical refinement and informed decision-making. These insights are essential for emergency managers, enabling more transparent deployment of AI systems, better communication with stakeholders, and the validation of assumptions derived from domain expertise.

4.1.6 How the FORESEE Should Be Used

Although the framework is organized into five conceptual stages (A–E), its practical use is not strictly linear. Instead, it reflects a workflow in which certain steps must occur first, while others can be revisited iteratively as new patterns emerge during the analysis. The purpose of this design is to allow the user to refine the modeling pipeline based on insights extracted from the data. Figure 18 demonstrates this process.

The process formally begins with Stage A, since no modeling or segmentation can occur without defining the geographic scope and assembling the relevant datasets. However, once the initial dataset is available, Stage D can be performed immediately and in parallel. Exploratory analyses, such as decomposition, autocorrelation structures, anomaly inspection, or stationarity tests, are frequently required early to reveal the intrinsic temporal behavior of the emergency events. These insights can guide decisions about temporal granularity (daily, weekly, hourly), preprocessing techniques, or whether more refined filtering is required. Thus, Stages A and D form an iterative exploratory loop, in which the dataset may need to be revisited, expanded, or cleaned based on statistical findings.

Stage B naturally follows after Stage A, because spatial segmentation depends on the presence of georeferenced data. However, regionalization also benefits from the insights obtained in Stage D. For example, if time series analysis reveals highly irregular patterns in certain months or years, or if spatial density maps indicate heterogeneous concentration zones, the user may choose to adjust the clustering strategy or number of clusters. For this reason, Stages B and D also operate iteratively, allowing for refinement of the spatial partitions based on the temporal signatures uncovered earlier.

Once regionalization is finalized and regions $c_0, c_1, \dots, c_n$ are established, the workflow proceeds to Stage C. This stage always occurs after B, because multimodal data depends directly on knowing which region each event and external covariate belongs to. Without spatial segmentation, it would be impossible to align meteorological, traffic, or calendar data with each localized time series. Stage C also includes the preparation required for multi-step forecasting, as well as the construction of model-ready sequences integrating observed values, static descriptors, and future-known features.

Following model construction, Stage E must occur after Stage C, because it depends entirely on the predictions generated by the model. However, Stages C and E also form an iterative cycle. The evaluation step may reveal the need to adjust hyperparameters, modify the multi-step forecasting strategy (e.g., recursive vs. direct forecasting), retrain the model with different feature sets, or reconsider the choice of architecture (e.g., comparing the performance of LSTM and TFT).

In summary, while the framework begins structurally with Stage A, the remaining stages interact dynamically, allowing cyclical refinement based on empirical evidence. This flexibility ensures that the resulting predictive models are not only technically sound but also deeply adapted to the local characteristics of the data and operational needs.

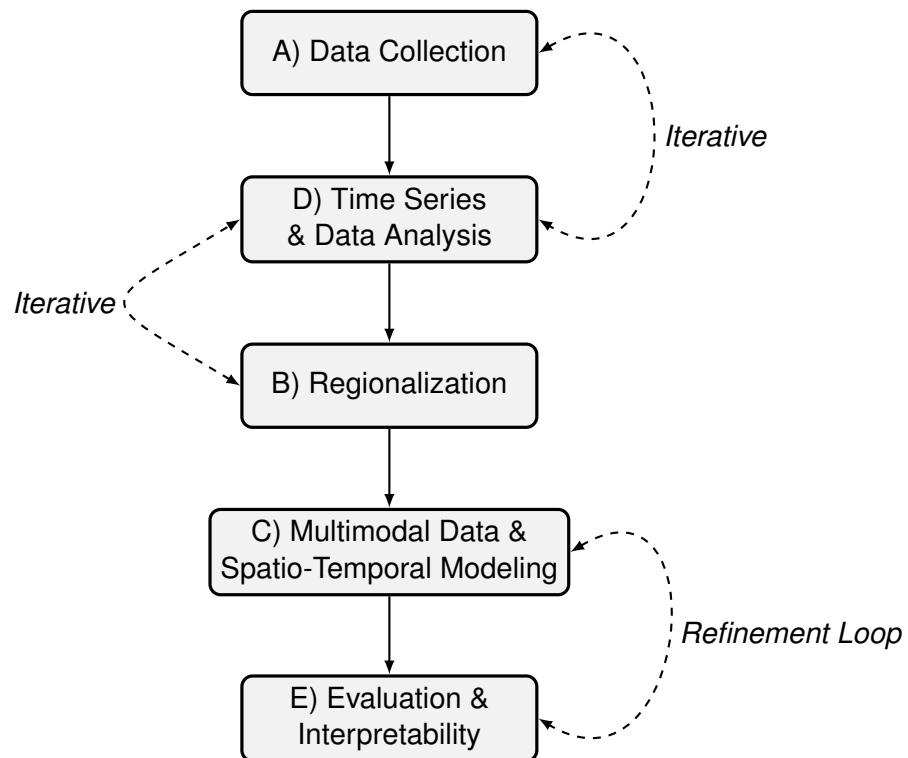


Figure 18 – Workflow of the FORESEE framework, showing the main pipeline ($A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$) and the iterative loops between exploratory analysis, regional modeling, multimodal Data, and evaluation.

5 CASE STUDIES

This chapter presents the results found by applying the proposed architecture. We present the case studies to demonstrate the generality of our framework and to gather evidence for the investigation. The first set of experiments focuses on confirming two foundational aspects of the framework: (1) the effectiveness of regionalized modeling through spatial segmentation, and (2) the capability of multi-step forecasting for predicting future emergency events across several temporal horizons. The second set of experiments evaluates two additional pillars of the framework: (1) the integration of multimodal data, such as weather and calendar variables, and (2) the interpretability of the predictive models. The experiments and a brief data description can be accessed upon request.¹

5.1 CASE STUDY 1: PREDICTION OF EMERGENCY EVENTS IN THE CITY OF FLORIANÓPOLIS

This section presents a case study proposed to test some methods and tools concerning the problem. We evaluate the space-time behavior of the problem by segmenting the regions in space and using an LSTM to measure the capability of prediction. This case study specifically confirms the steps A), B), and certain portions of C) and D) (see to Figure 19). The results presented in this case study is partial or fully presented in (VIEIRA ROQUE et al., 2025).

¹ Available in: <https://github.com/Felipe-Roque/OcurrencesPrediction/blob/master/cap5.ipynb>

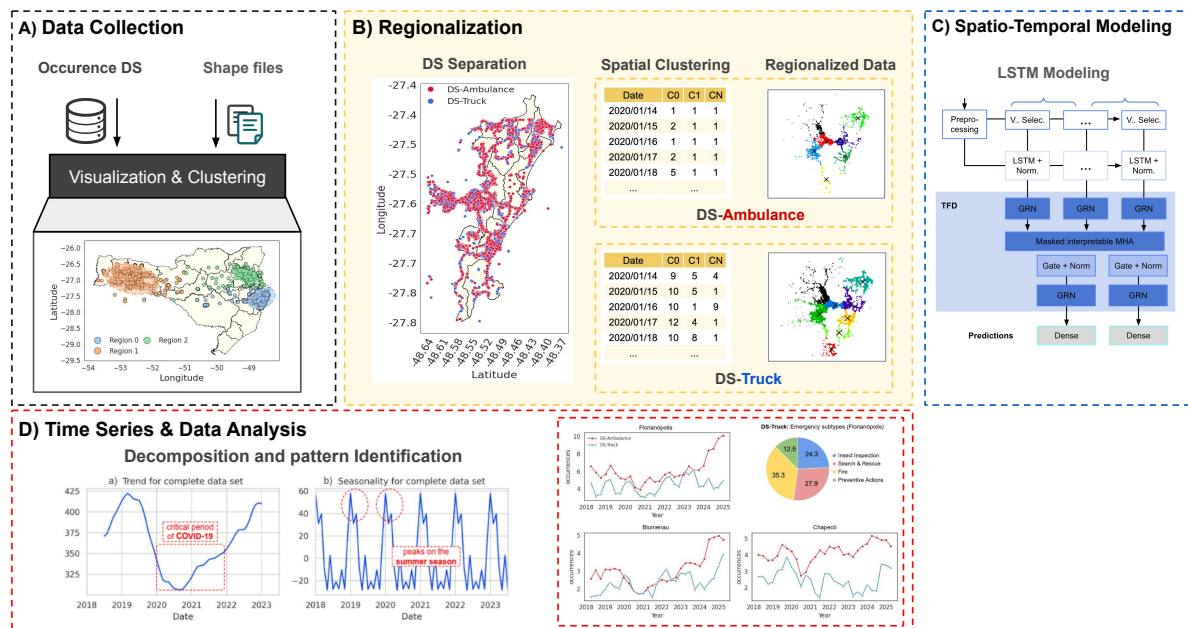


Figure 19 – Framework components applied in Case Study 1.

5.1.1 Data Collection

The present study focuses on predicting emergency events in Florianópolis, Brazil, a city with a population exceeding 500,000 inhabitants and multiple military fire stations spread across its territory, with 675.4 km². These stations respond to various incidents, including fires, search and rescue operations, floods, and even pre-hospital care. The data is categorized into four distinct emergency event types: pre-hospital care, fire, search and rescue, and dangerous products, as each type presents unique characteristics and frequency patterns.

The data consists of real occurrences from the Military Fire Corps (MFC) of the State of Santa Catarina, Brazil, which is generated by an automated system that tracks the progress of each event. When an individual contacts the emergency services, the system is activated, generating an occurrence that includes the location of the event and the activation of the closest emergency station.

Table 6 illustrates that each occurrence is associated with a unique identifier (ID), a date, and geographical coordinates. Additionally, the type of emergency event is also recorded, with pre-hospital care being the most common occurrence.

Table 6 – Real occurrence data.

ID	Date	Longitude	Latitude	Type
2020001	01/01/2020	-48497971	-27546850	Fire
2020233	21/01/2020	-48497967	-27546857	Dangerous Products
2020456	11/01/2020	-48497956	-27546789	Pre-hospital care
2020249	25/01/2020	-48497890	-27546798	Search and Rescue
2020598	29/01/2020	-48497789	-27546878	Dangerous Products
2020489	12/01/2020	-48497876	-27546760	Fire

The original database consisted of aggregated records of all events that took place in various cities of Santa Catarina. However, we only evaluated the occurrences that took place in Florianópolis. It is important to note that the initial database contained duplicate records since the same occurrence was represented with different numbers of vehicles and people involved. To conduct our analysis, we segregated the pre-hospital care events into one group and another large group that included fire, hazardous products, and search and rescue services. This decision was based on the equipment available with the service organizations. Typically, there is an ambulance for pre-hospital care and a fire truck to handle other incidents. However, this does not imply that the two cannot be included in the same emergency event.

Handling experiments, we encountered data collection issues, particularly in the recording process within the relational database. Originating from an actual emergency system, certain incidents lacked precise location details but included information about the neighborhood and complete address. To solve this problem, we performed imputation, assigning the same location as another occurrence with matching neighborhood information, allowing for the utilization of this dataset.

Following the data cleaning and selection process, the entire database was categorized into two distinct categories, namely DS-Ambulance occurrences and DS-Truck occurrences, comprising 24806 and 8053 observations, respectively. Duplicated and missing data were eliminated to ensure an accurate representation of the problem.

We aggregated the series by day since the daily forecast is relevant to the problem. Table 7 provides a comprehensive overview of the data for the two sets evaluated. After all the treatments to extract the relevant information for the models and Times Series analysis, the occurrence data are represented by Figure 20. Each large block in the figure represents a year of analysis. The rows represent the days of the week and the columns the weeks distributed by month. The redder the color, the more occurrences on the day there were. The year of 2019 has the highest number of occurrences (7). The year 2023 was only evaluated in half, accounting for half of

the occurrences. However, from Figure 20 we can see that this year's occurrences compare to 2019 in intensity. We explore these behaviors further in the time series analysis below.

Table 7 – Data description.

	DS-Ambulance	DS-Truck
Total Number of Days	1943	1943
Total Number of Events	24806	8053
Total Average by day	10.96	2.2
Total Standard Deviation	4.55	3.59
Total Events 2018	4522	1727
Total Events 2019	4991	1823
Total Events 2020	3698	1543
Total Events 2021	4103	1246
Total Events 2022	4525	1071
Total Events 2023	2967	643

5.1.2 Time Series Analysis

We analyze all data together (before segmentation by k-means), intending to identify patterns for the later relationship with other results. In this analysis (see Figure 22), we can see the components decomposed for the city of Florianópolis, observing different periods. The trend is the value that causes increasing or decreasing patterns in a time series. Seasonality is a cyclic event that occurs in a time series for a short time and causes increasing or decreasing patterns for a short time in a time series. Residual are the random variations in the time series. In another way, Autocorrelation analysis helps to find good relationships in past terms on data. The next sections show this analysis.

5.1.2.1 Autocorrelation analysis

In the time series evaluation in this work, the test confirmed the stationary behavior. Figure 21 shows the ACF plot and PACF plot for the DS-ambulance.

One of the initial objectives of this analysis was to find an adequate value for the prediction window (Look-back parameter of the sliding window). As we can see from the analysis of figure 21, the initial days seem to be good values for the prediction window (days in the sequence of the current day). The first three days are the ones with high values of ACF and PACF. The dashed blue lines indicate whether the correlations are significantly different from zero (white noise series).

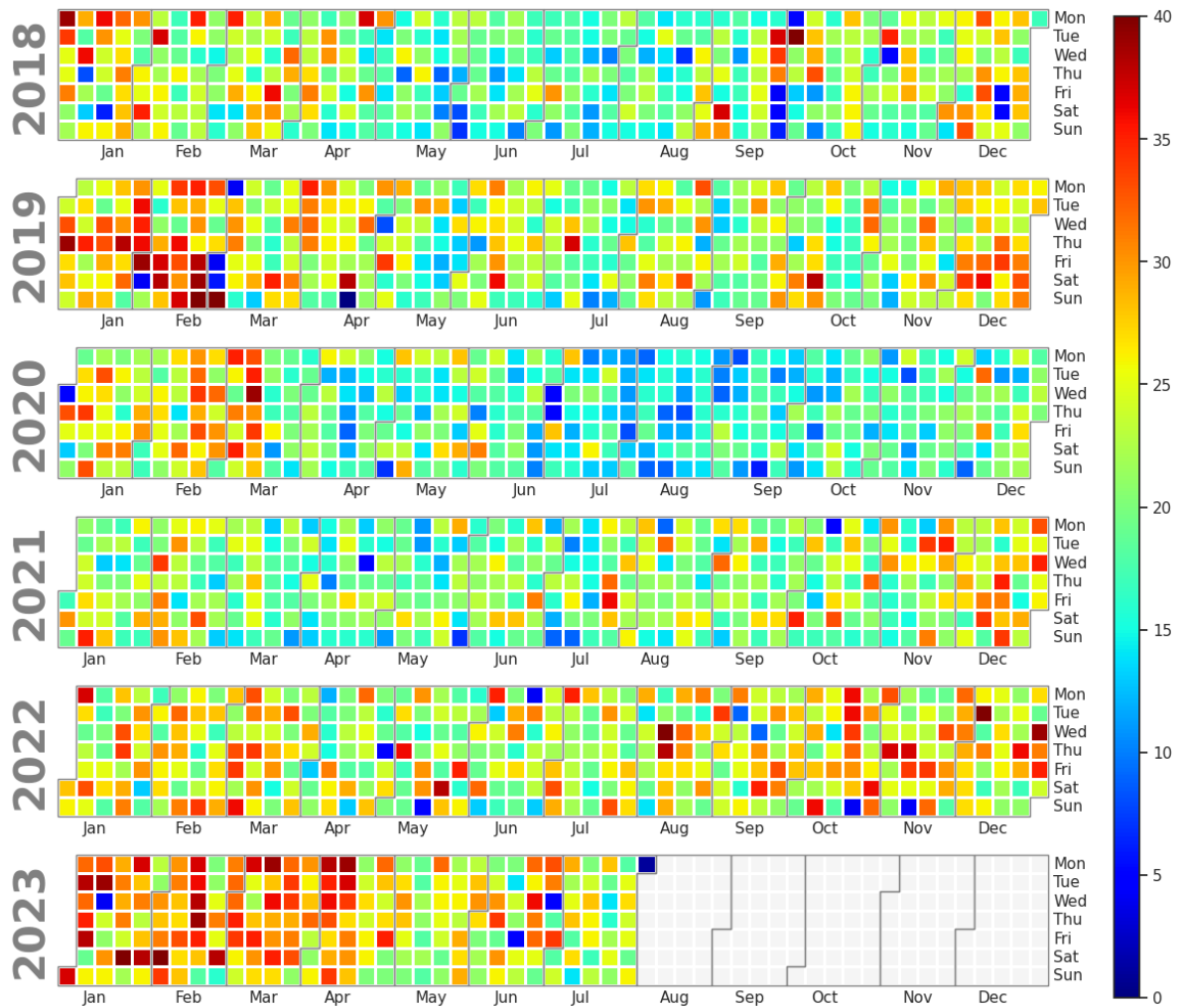


Figure 20 – A representation of occurrences by day of the week and year.

We focused this analysis on the last 300 days (spanning the latter half of 2022 and the first half of 2023), as this period reflects a return to normality in emergency occurrences following the disruptions caused by COVID-19. However, although it is relevant, we run an optimization technique to find the best value for the problem (see Section 5 how these values can be found).

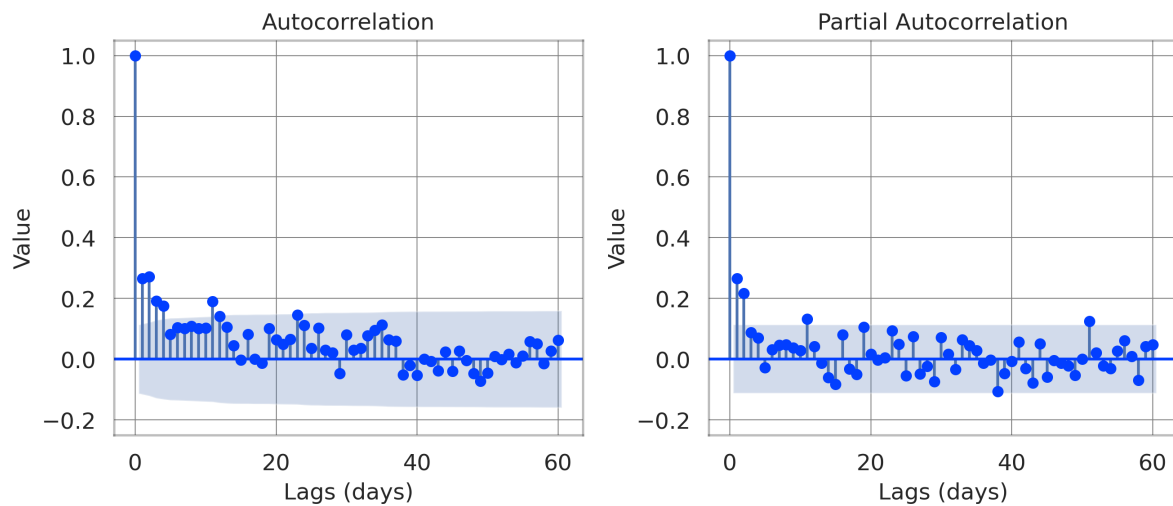


Figure 21 – ACF and PACF plots for pre-hospital occurrences (DS-ambulance).

5.1.2.2 Times series decomposition

Another way of evaluating the time series can be seen in Figure 22. A time series can be decomposed into smaller signals. This decomposition can facilitate analysis and create better insights into the data. With decomposition, we can analyze the parts of the series separately. In addition, we can analyze different periods in search of a better understanding of the data.

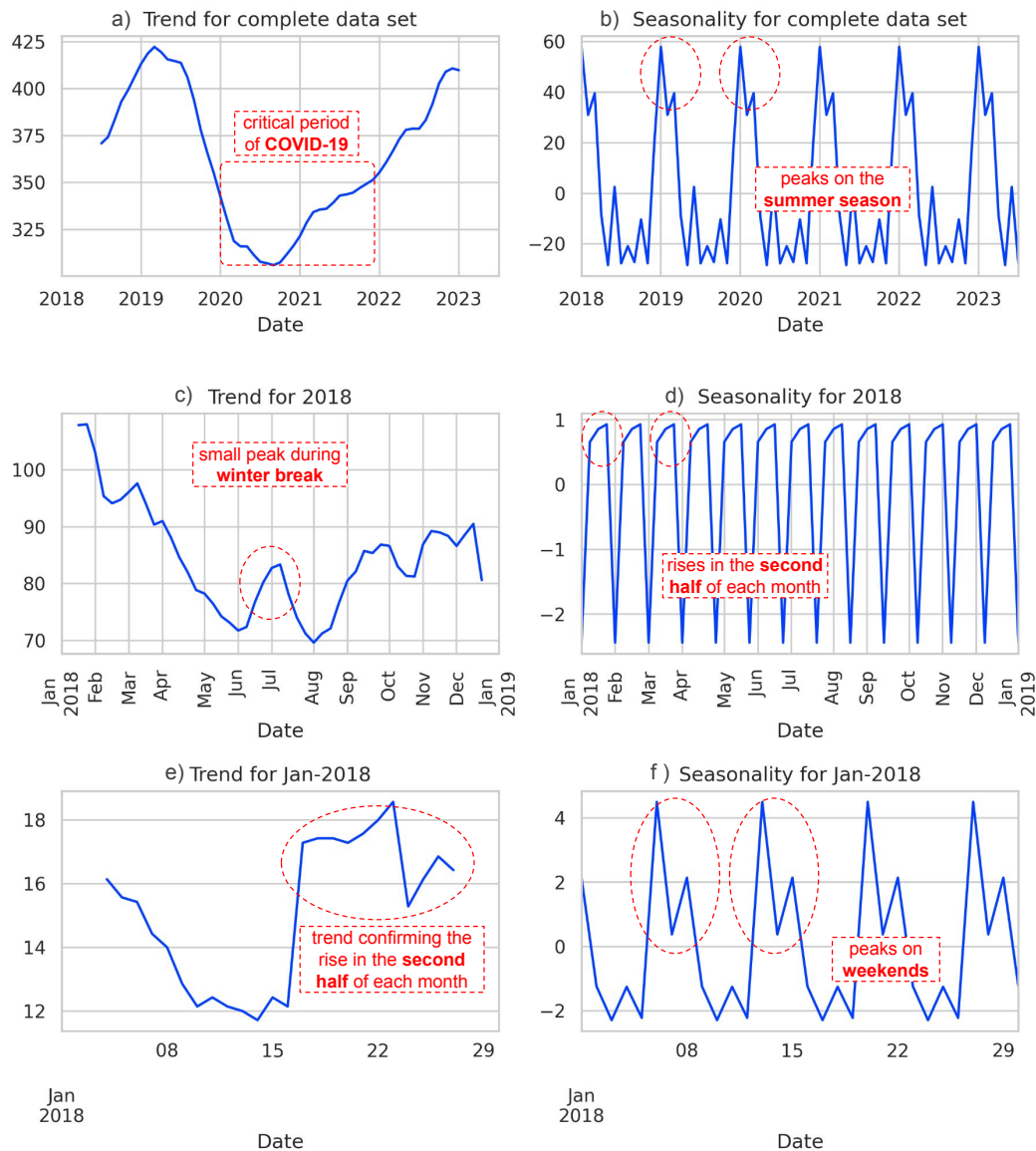


Figure 22 – Decomposition of the series into trend and seasonal for all data.

For example, in Figure 22, we see the trend (Figure 22a) component, and we can infer some behaviors in the analyzed series. We can see that initially, the time series had a strong upward trend around January 2019. Between that month and the beginning of 2020, we can see the period of the COVID-19 pandemic, which significantly reduced the average number of daily occurrences. This behavior can also be seen in Figure 20, where the intensity of occurrences decreased. After this pandemic, the occurrences gradually returned to the normal pattern, with an upward trend between 2021 and 2023.

In Fig.22e, we can see the image representing the seasonality decomposed into 1 month. For this analysis, we set a periodicity of 7 days, showing the rise in the second half of each month. This pattern is confirmed by the image in the middle right.

There is an interesting behavior on weekends in this data, (Fig. 22f). More specifically, in the days leading up to the weekend, on Thursday and Friday, and on Saturdays. On these days, we can observe an increase in the number of occurrences compared to other days of the week, as shown in the Figure. These days, there are usually the biggest parties, and when people get together more. In Figure 22c, we can see an indication of the peak during the winter break in 2018.

Figure 22b is the seasonality component over the years. It represents the seasonality, evaluating 30 days. The objective is to find a pattern referring to the months of the year or the seasons of the year, for example. We can see that we have increased the number of occurrences in the summer months. This observation can also be confirmed by Figure 20, where the color intensity is redder between December and March when summer occurs in the Southern Hemisphere. This increase is caused by the movement of people to the city, as Florianópolis is a national tourist destination.

Across all decompositions, the residual component remains within expected bounds when compared to the trend and seasonal structures extracted from the data. In the full-year analysis, the magnitude of the residuals is indeed greater than the seasonal component, but it remains consistently smaller than the trend component, indicating that the long-term behavior of the series dominates the observed fluctuations. This pattern is also confirmed when extending the analysis to all years in the dataset, showing that irregular variations do not exceed the underlying long-term dynamics. An interesting observation emerges from the January-only decomposition: both the seasonal and trend components are individually larger than the residual component, suggesting that short-term behavior in this month is more strongly governed by structured, predictable patterns. This reinforces the view that, although unpredictable variations exist, the major drivers of the series are captured by the trend and seasonality. Figure 23 presents these results.

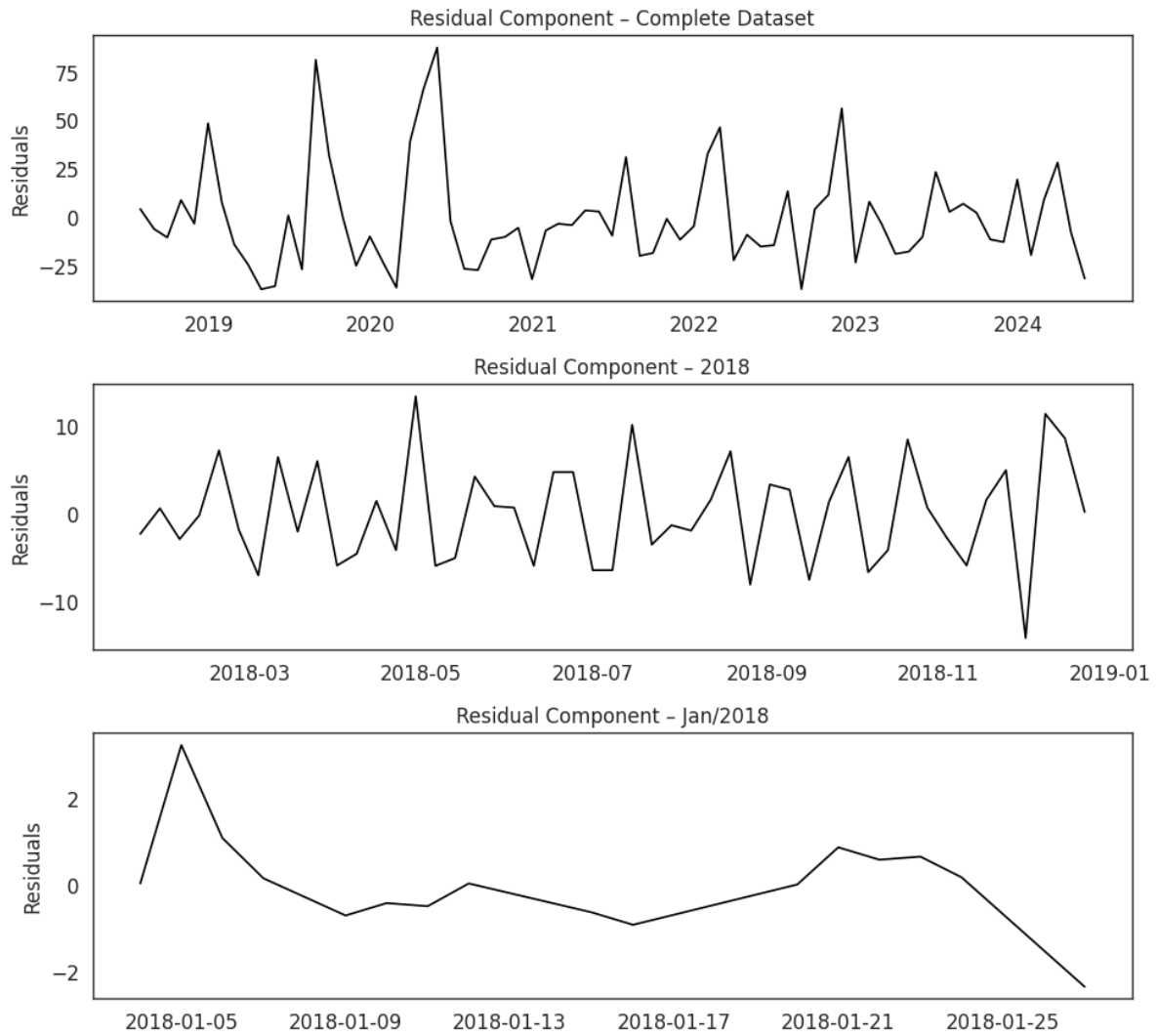


Figure 23 – Residual Component.

5.1.3 Regionalization

For each of the datasets (DS-Ambulance and DS-Truck) we perform the segmentation, in this way, we can separate the two behaviors of the types of occurrence. Figure 24 illustrates how we employ the step of the framework for this problem with two distinct datasets.

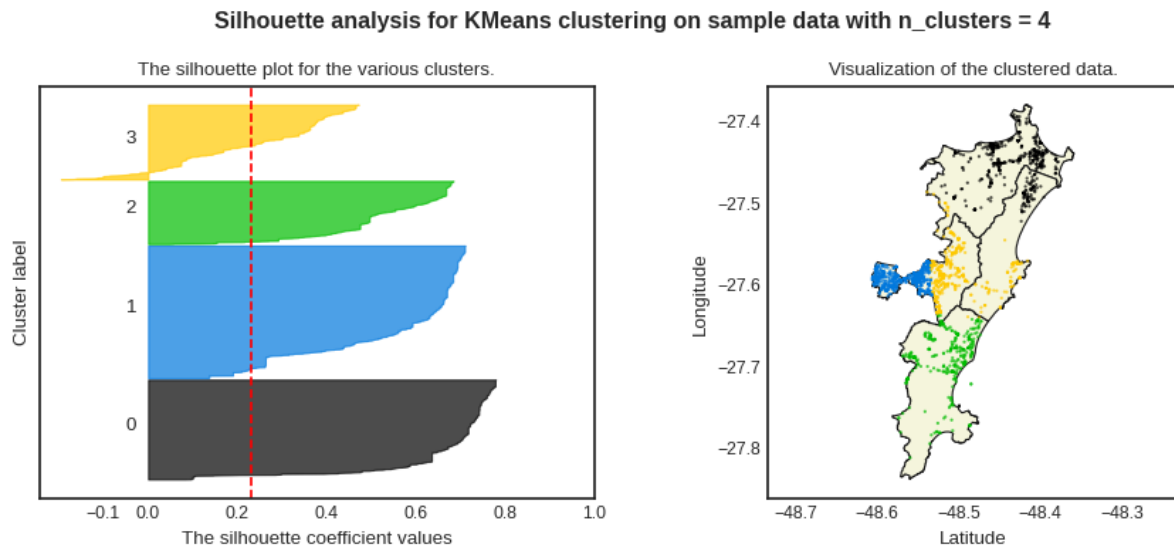


Figure 25 – Best k value for the problem and its visual separation seem from silhouettes for DS-Truck.

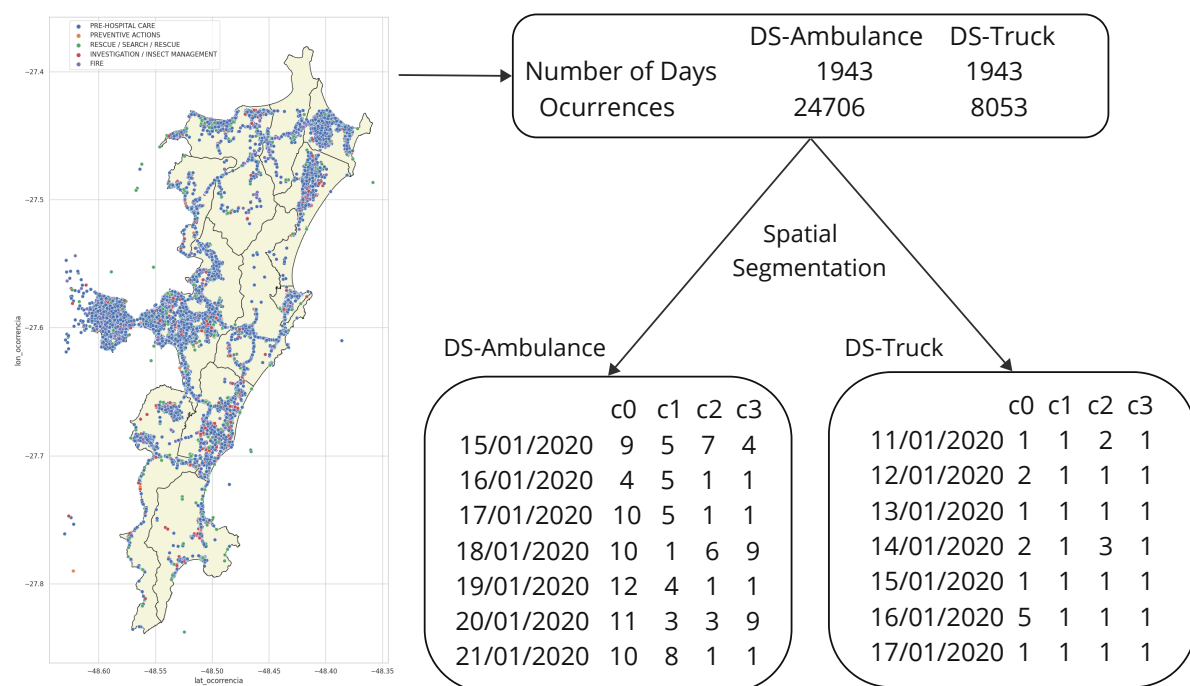


Figure 24 – Visual illustration of the data preparation methodology.

For this work, we evaluated 8 different values of groups to define what best separates the observations in the problem. We defined the number 4 as the smallest possible value for this analysis because the fire department has four emergency dispatchers spread across Florianópolis, Table 8 presents this result. The idea of testing

higher values is that this could indicate the need to build more stations if this improves the analysis and the efficiency of rescue operations. In our Silhouettes' evaluation, the best value of groups was 4 for k-means and k-Medoids. For Agglomerative the best value was 5. The results indicate that dividing the city into four regions leads to a more homogeneous allocation.

Table 8 – Number of clusters per average value of silhouette.

Number of clusters	Avg sil for DS-Ambulance	Avg sil DS-Truck
k = 4	0.55	0.50
k = 5	0.49	0.47
k = 6	0.46	0.44
k = 7	0.47	0.43
k = 8	0.49	0.45
k = 9	0.49	0.47
k = 10	0.50	0.47
k = 11	0.50	0.47

As can be seen in table 35, the silhouettes values are very close. As the value of k increases, the average of the silhouettes value tends to get closer to 0.5. This analysis can be useful in deciding where and how to install new emergency response sites. We can also observe that the best value of k for both problems was the same. The distribution of emergency stations is the same for the datasets, however, the dynamics of occurrences are different between them. It is relevant to verify this situation for decision-making to create or relocate emergency teams.

With the definition of $k = 4$ for the segmentation of locations, the initial series was divided into four series for DS-Ambulance and DS-Truck events, defined by location (see in Figure 25). The figure illustrates that k-means is effective for clustering observations based on geographical coordinates (Longitude and Latitude). The city has a different population distribution throughout its area. There are denser regions, like where we can see the green dots. The other three regions are less populated, however they also have an urbanized area, amidst areas of native forests.

For the DS-Ambulance, the figure representing the segmentation by k-means was very similar to Figure 25. In this way, we prefer not to insert it in this work. It is important to point out that the incidents attended by the fire truck tend to be concentrated in the same region as the ambulance incidents, in areas with a concentration of population.

Figures 26 represent the series over time with the smoothing of the quarter for the DS-ambulance dataset. A similar behavior is encountered in DS-Truck dataset.

The k-means segmentation was done to extract the specific behavior for each location. Knowing that there are different teams for different locations in the city.

We can observe that one of the clusters (c2), in Figure 26, has, on average, more observations over time, Table 9. The others have very similar behavior to each other. This can be explained by the team's service region (the location of the rescue team), where c2 is a region with a greater concentration of people than the other regions.

On the other hand, the DS-Truck data have values closer to each other, in the 30-day moving average, not exceeding 10 occurrences per day. This indicates that occurrences of this type are less frequent than those observed in the DS-Ambulance. In clusters 1, 2, and 3, we have low values, close to 0, for truck occurrences. This type of information can be relevant to the organization of emergency teams.

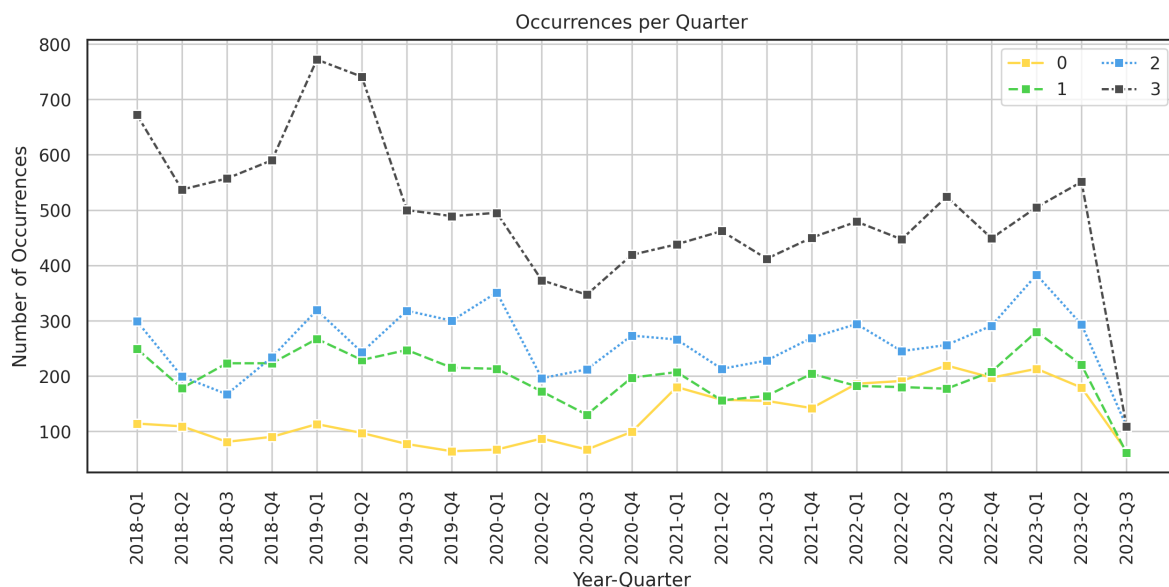


Figure 26 – Four clusters for DS-Ambulance events over time divided by the clustering technique with moving average per quarter.

Table 9 – Description of cluster data.

	DS-Ambulance			DS-Truck		
	Total events	Mean	std	Total events	Mean	std
c0	2348	1.44	1.34	1788	0.92	1.03
c1	5958	2.92	1.80	3173	1.63	1.44
c2	11311	5.55	3.47	1602	0.82	1.02
c3	4582	2.24	1.59	1490	0.76	0.98

5.1.4 Spatio-Temporal Modeling

This section presents the results obtained using our custom-designed, optimized LSTM architecture. We also compare with a classic time series prediction and machine learning techniques. Furthermore, we present two approaches to predict days ahead with the LSTM model.

The experiments were conducted on a system running Ubuntu 22.04, equipped with an Intel 13th Generation Core i5-13420H processor featuring eight cores and a clock speed of up to 4.6 GHz. The system also included an NVIDIA GeForce RTX 3050 GPU with 6GB of dedicated memory, 16GB of DDR5-5600 MHz RAM, and a 512GB Solid State Drive (SSD) for storage.

5.1.4.1 LSTM for predicting emergency events

The model underwent training for 500 epochs, and training performance was assessed using the Least Absolute Deviations - L1 loss metric. Adam Optimizer is employed to control the learning rate and minimize average loss. Additionally, we implemented Time Series cross-validation with five folds for data to avoid overfitting. We splitted the data in 70% for training, 20% for validation, and 10% for testing. Simultaneously, Optuna, a hyperparameter optimization tool was utilized to adjust the model parameters. We look for loss minimization using TPE. TPE works by building probabilistic models that estimate the performance of different hyperparameter configurations. Unlike traditional grid search or random search, TPE focuses on exploring the most promising regions of the search space based on past evaluations, making it more efficient for complex models with numerous hyperparameters. Shifting our focus towards data-centric enhancements, our approach aimed at improving overall performance. The outcomes are visually depicted in Figure 27.

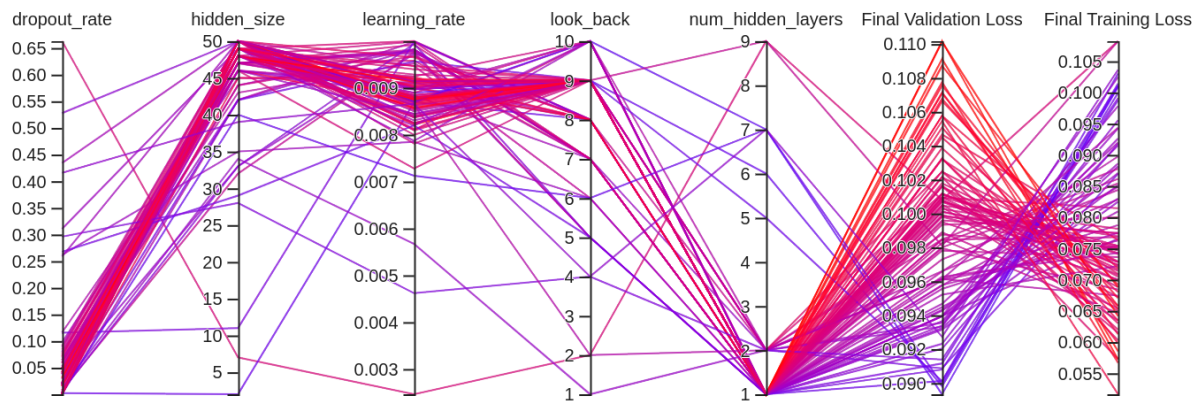


Figure 27 – Runs optimization for DS-Ambulance occurrences.

We monitor the validation loss in the optimization phase to prevent the model from overfitting, focusing on key parameters such as hidden size (hs), representing the number of neurons in the hidden layer; lb, indicating the number of time steps considered for predictions and influencing the model's temporal pattern understanding; learning rate (lr), controlling the step size in optimization and affecting the model's adaptability; dropout rate (dr), a regularization technique that randomly omits a fraction of units during each epoch to prevent overfitting; and the number of hidden layers (nhl), defining the neural network's depth and impacting its capacity to learn intricate features. This targeted optimization strategy aims to improve the overall model performance, guaranteeing an optimal configuration for the specified parameters. To achieve this, we conduct 350 trials for each dataset to identify the most effective parameters for the model.

Examining Figures 27, it is evident that certain parameters lead to overfitting of the model. Numerous parameter combinations, particularly those with a hs near 50, exhibit training loss surpassing the validation loss. Similarly, the optimization process tends toward higher lb values, suggesting that a more extended historical perspective correlates with improved error outcomes. The optimal configuration for the nhl was near 1, too. The best parameter configuration for the model was hs 48, lb 10, lr 0.009, dr 0.002, and nhl 1.

We also observed the impact of key parameters on model performance. Changes in hs, for example, directly influenced the model's capacity to capture complex patterns, with larger hidden sizes (around 50) improving performance up to a point before overfitting began. Similarly, the lb parameter significantly impacted the model's ability to understand temporal trends, with longer periods typically leading to more accurate predictions, as shown in Figure 27. Models also present sensitivity to the lr parameter,

with lower rates resulting in the model missing optimal configurations. dr adjustments played a critical role in regularization, with lower values giving the best results in validation loss. We could also observe that low nhl gives the best results for this problem.

To quantify the relationships between the MAE and the model parameters, we performed a Pearson correlation analysis, presented in Table 10. The analysis highlights the importance of parameter tuning: a positive correlation indicates that increasing the parameter value is generally associated with an increase in MAE, suggesting a need for careful adjustment to avoid negative impacts on model performance. Conversely, a negative correlation suggests that increasing the parameter value tends to reduce MAE, helping to mitigate overfitting. This analysis highlights the importance of specific parameter adjustments, suggesting that reducing the hs due to its positive correlation with MAE and increasing the nhl due to its negative correlation can effectively improve model performance for DS-Ambulance.

Figure 28 illustrates the gradual decrease in loss over epochs for DS-Ambulance. We depict the distinct patterns between validation and training data to assess potential overfitting. For DS-Truck dataset model, we have similar behavior to the one described in Figure 28, so it was omitted for better visualization. Figure 27 shows that some parameter combinations lead to overfitting, emphasizing the significance of scrutinizing training and validation data.

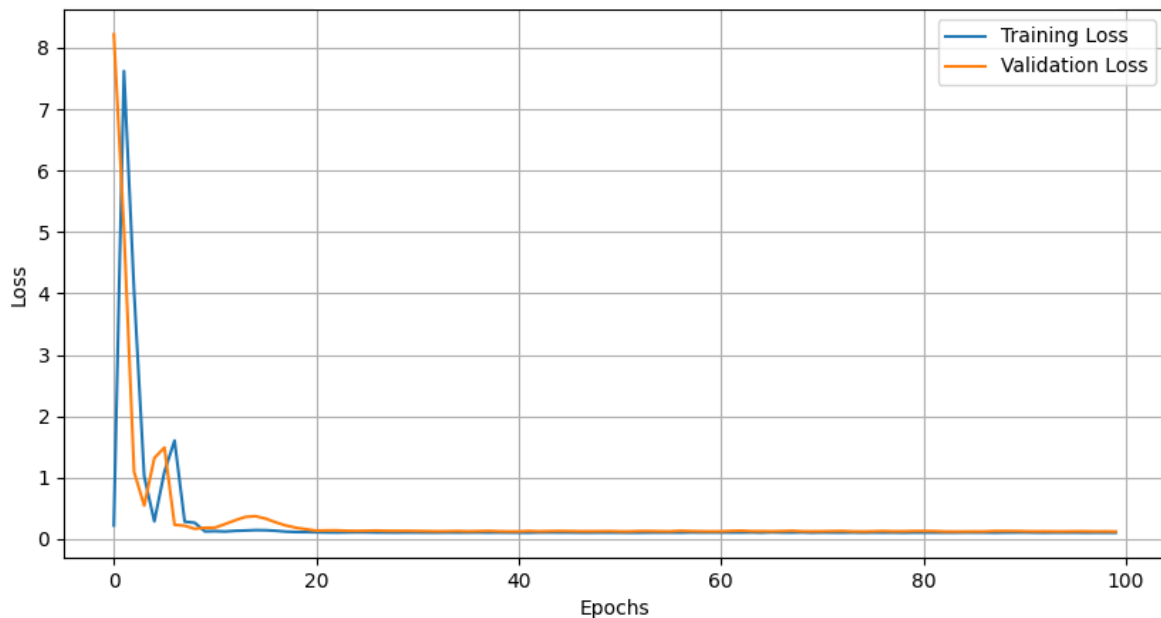


Figure 28 – Loss plot for Ambulance occurrences.

Table 10 – Pearson correlation of tested parameters with the Mean Absolute Error.

	hs	lb	lr	dr	nhl
DS-Ambulance values	0.39	0.32	0.04	-0.28	-0.31
DS-Truck values	0.45	0.09	-0.46	-0.33	-0.43

5.1.4.2 LSTM performance comparison with different models

The experimental designs of traditional TS models for regression are first explained. To select the best smoothing order for MA, an analysis with the autocorrelation function to determine when it becomes closest to zero value was performed. The best value of the smoothing order was set to 10 because it was near the average. To find the best parameters of the ARIMA method, we use the *autoarima* (SMITH et al., 2017) package. In the STL approach, we optimized the period that refers to the length of the seasonal component in the decomposition. For DS-Ambulance, we identified the optimal parameters as (1,0,1), where the first term p signifies the autoregressive term, the second term q denotes the differentiation term, and the third term d represents the moving average term of the *autoarima* model. Similarly, for DS-Truck, the (1,0,1) model emerged as the best, determined by the Akaike Information Criterion (AIC) score in the ARIMA model comparison. Table 11 shows the results for different regression models.

To find the configuration with the best parameters of the machine learning techniques, the exhaustive grid search technique was also utilized. As a result, in the case of SVM, the penalty parameter was set to $c = 0.1$ and ϵ for insensitive loss. Since DT is a non-parametric supervised learning technique, we scrutinized the maximum depth (md) parameter. In the case of MLP, we assessed the Activation Function (af) and hidden layer size (h/s). Meanwhile, the LSTM parameters were optimized as explained in the preceding sections. The results are detailed in Table 12.

One way to evaluate the performance of a regression algorithm is to use error functions. For this study, we used 3 functions. The MAE, the MSE, and the Mean Squared Error (RMSE). We use these metrics to evaluate the model's performance after the training phase.

Table 11 – Results of Emergency Event's regression models.

Models	DS-Ambulance			DS-Truck		
	MAE	MSE	RMSE	MAE	MSE	RMSE
ARIMA	3.22	16.65	4.08	2.01	5.67	2.38
STL	4.32	29.39	5.42	1.95	6.24	2.49
MA	2.78	11.99	3.46	1.43	3.39	1.84
DT	3.85	23.90	4.88	1.60	4.68	2.16
MLP	3.30	17.56	4.19	2.10	5.95	2.44
SVM	3.41	18.81	4.33	1.62	4.33	2.08
LSTM	1.29	3.02	1.73	0.71	1.03	1.01
REF-LSTM	1.57	4.03	2.0	0.75	1.13	1.06

As shown in Table 11, the LSTM model outperforms all proposed approaches. To provide a consistent baseline for comparison, we also evaluate a reference LSTM model, denoted as REF-LSTM, whose architecture and hyperparameters were fixed according to the configuration proposed by Cortez et al. (2018). The is also concerned with predicting daily events. The LSTM proposed in our work is superior to this approach for both datasets (DS-Ambulance and DS-Truck).

Table 12 – Parameters optimization for different approaches.

Model	Parameters	DS-Ambulance	DS-Truck
ARIMA	p, q, d	$p=1, q=0, d=1$	$p=1, q=0, d=1$
STL	period	90	90
MA	SO	10	10
DT	md	1	2
MLP	af, hls	tanh, 100	tanh, 50
SVM	c, ϵ	$c=0.1, \epsilon=0.1$	$c=0.1, \epsilon=0.2$
LSTM	lb, lr, dr, hs, nhl	10, 0.012, 0.5, 48, 1	8, 0.013, 0.5, 50, 1
REF-LSTM	Fixed	Fixed	Fixed

5.1.4.3 Comparing look-ahead LSTM scenarios

To assess the predictive capabilities of our proposed approach, we conducted experiments to predict several future days using the lstm model. Using the optimal parameters identified for the lstm model in table 12, we produced forecasts that extended up to 10 days. We analyze two different methodologies to solve the problem of multi-step time series prediction (TAIEB; HYNDMAN, et al., 2012): (i) **direct forecasting**, which involves training a different model for each step ahead; and (ii) **recursive forecasting**, in which the model output at time t is reused as input to slide the prediction horizon one step ahead to predict $t + 1$ using the same model.

Table 13 – Comparing lstm regression results for different look ahead scenarios.

Steps Ahead	DS-Ambulance		DS-Truck	
	mae - Recursive	mae - Direct	mae - Recursive	mae - Direct
1	1.29	1.29	0.71	0.71
2	1.59	1.68	0.73	0.79
3	1.59	1.60	0.72	0.75
4	1.58	1.69	0.72	0.78
5	1.58	1.61	0.72	0.77
6	1.57	1.66	0.71	0.77
7	1.57	1.68	0.71	0.78
8	1.57	1.64	0.71	0.77
9	1.57	1.61	0.71	0.76
10	1.57	1.63	0.71	0.75

Table 13 presents the results for each of the strategies (full results including MSE and RMSE are presented in Appendix B.1.2) adopted to predict days ahead. As we can see, for both approaches, the best strategy would be to use one day ahead. This is expected as the model was optimized for the look-ahead of 1. We limited the approach to 10 days since the analysis of pacf and acf indicates that after this value, the data no longer has autocorrelation 21.

Another interesting observation is the superiority of the approach where the input is fed back to the lstm output compared to retraining the model for multiple look-ahead windows. This superiority can be attributed, in part, to the model being optimized for a single day. Furthermore, it's noteworthy that the model's prediction output undergoes minimal changes with each prediction for $(t + 1)$. Consequently, small alterations in the input have almost no impact on the error. This consistent behavior is observed across both datasets analyzed (DS-Ambulance and DS-Truck). Figure 29 compares mae error in both datasets.

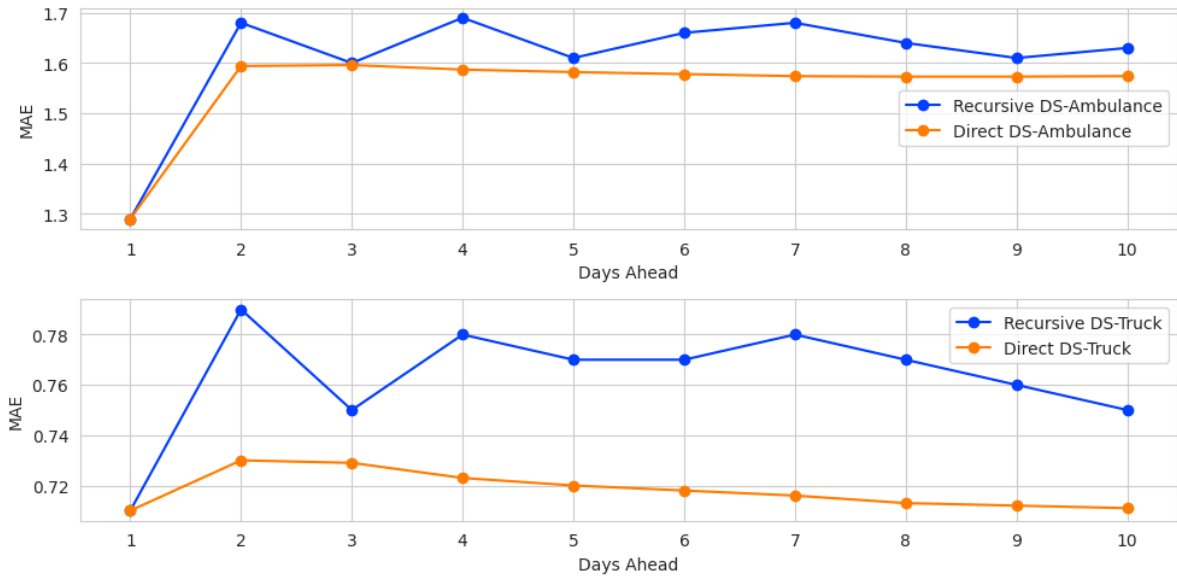


Figure 29 – Error for different approaches.

5.1.4.4 Model and framework generalization analysis

To assess how our model and our framework generalize to different data distributions, we tested our LSTM-based approach considering two scenarios: (i) with LSTM model retraining, whose purpose is to validate that our framework can be easily reproduced to different cities/regions; and (ii) using the same LSTM model from the Florianópolis with no retraining or fine-tuning, which aims at evaluating how our LSTM model generalizes as well. We have selected data from the city of Itajaí, also located in the state of Santa Catarina.

In the first scenario, the entire pipeline was applied to this new dataset (DS-ITA), which was split into a 70% training set and 30% for testing, using both the ambulance and fire truck events as a target. For the DS-Ambulance dataset, the model parameters that were selected using our framework are the following: h_s of 32, a l_r of 0.00763, a d_r of 0.07408, a l_b of 9, and 2 n_{hl} for the DS-Ambulance dataset. For the DS-Truck dataset, a different model was employed with the following parameters: a h_s of 19, a l_r of 0.000148, a d_r of 0.45726, a l_b of 7, and 5 n_{hl} .

In the second scenario, the same model and test dataset were used for inference only. We fixed the number of region clusters to 4 in both scenarios to keep model compatibility and simplify comparisons. Table 14 presents the results.

Table 14 – LSTM error metrics for the city of Itajaí (DS-ITA).

Scenario	DS-Ambulance			DS-Truck		
	MAE	MSE	RMSE	MAE	MSE	RMSE
With LSTM retraining	0.96	4.95	2.22	0.34	0.45	0.67
Without retraining	2.27	19.21	4.38	0.53	1.00	1.00

As we can see in Table 14, we have compared the retrained LSTM pipeline with posterior tests using data originating from DS-ITA. Additionally, we compared this approach with a model trained on data from Florianópolis and tested using data from Itajaí. The results show that the retrained model using training and test data from Itajaí performs better than the Florianópolis-trained model. This supports the conclusion that specialized models are required for this problem, as region-specific characteristics significantly influence prediction accuracy. This comparison highlights the adaptability of our approach to different datasets, demonstrating its efficacy across multiple space-time scenarios.

5.1.5 Discussion: A LSTM approach to predict emergency events using spatial features

The work of first responders can often be very difficult. The first response needs to be fast and effective. For this response to have these characteristics, teams must be well distributed over a certain region. The limited supply of personnel and tools may make this distribution challenging. To be able to evaluate and propose a prediction solution, we use some machine learning techniques and time series analysis. First, we filter the data, which at its origin was mixed among all types of occurrences, Table 15. We divided them into two main datasets, DS-Ambulance and DS-Truck, for a better understanding of each type of event.

From the initial time series, we decided to segment them into clusters, based on their location, Figure 25. This segmentation can help decide where to place a new dispatch station since the outflows were segmented based on the exact locations of the occurrences. Furthermore, in this work, we used the Silhouettes metric to identify the best value of k for the problem. We chose to use $k=4$ since, for the evaluated city, it seemed to be the best choice (there are 4 dispatch points in the city).

The analysis using Time Series made it possible to understand some patterns in the data. For example, for the analysis with the ACF (Figure 21), we can identify a good prediction window value for MA algorithm. Furthermore, with the decomposition, Figure 22 of the series into its trend, seasonality, and residual components. This

decomposition allowed us to understand the behavior of occurrences throughout the week and between the seasons. This result can help decision-makers in the allocation of resources and the decision of public policies for awareness, since emergencies can be identified in certain periods based on the decomposition. The decomposition also clearly showed the influence of the COVID-19 pandemic period on the pattern of occurrences. This analysis can extrapolate models during abnormalities caused by similar phenomena.

Finally, the optimized architecture LSTM model proved more accurate than all models evaluated for these problems, Table 11. We also proposed two methodologies to predict days (Table 13). Models capable of predicting the occurrence count several days in advance are crucial in optimizing resource allocation. This outcome facilitates decision-making grounded in metrics rather than intuition. In practical scenarios, the ability to swiftly reallocate teams in response to a predicted peak, as indicated by the prediction model, enhances operational efficiency.

5.2 CASE STUDY 2: TFT MODEL FOR MACRO-REGIONAL EXPERIMENTS

This case study focuses on forecasting emergency events across multiple cities in Santa Catarina, Brazil, using real-world multimodal data. The dataset includes time series of firefighting and pre-hospital care occurrences, enriched with meteorological and calendar-based variables. We segment the data spatially based on regional patterns and train separate models for each segment. The proposed approach employs a TFT model to perform multi-step forecasts, predicting daily emergencies several days ahead. This case study specifically confirms the steps A), B), C), D), and E) (see Figure30).

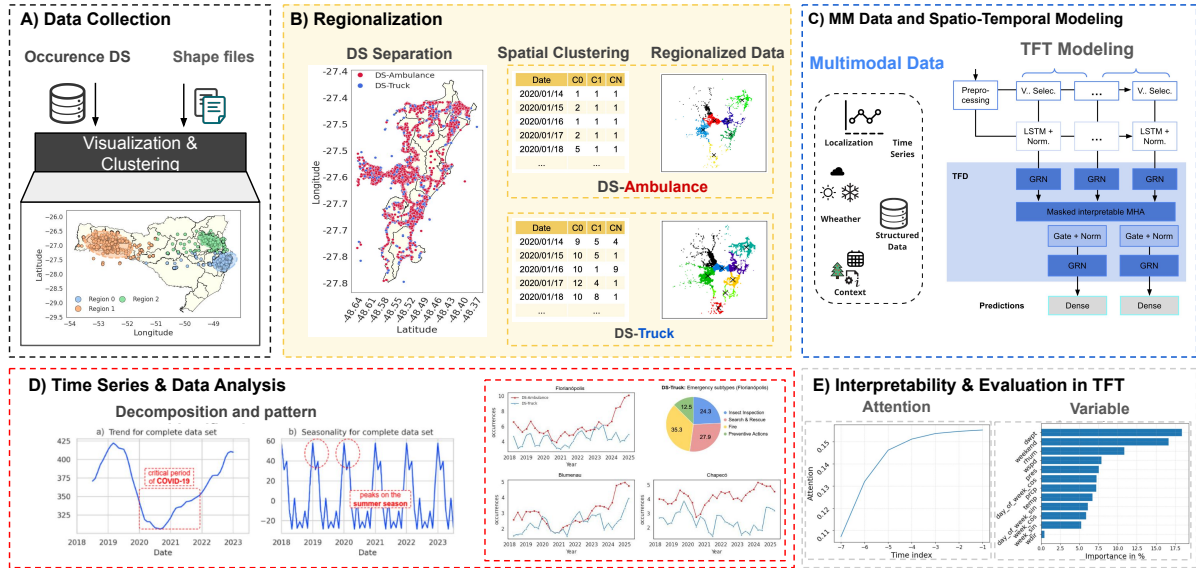


Figure 30 – Framework components applied in Case Study 2.

5.2.1 Data Collection

The dataset encompasses emergency occurrences from January 2018 to October 2024 across four cities Florianópolis, São José, Palhoça, and Biguaçu, forming a macro-region in Santa Catarina, Brazil. This region, referred to as Macro Flo, is serviced by emergency stations that address a diverse array of events, including fires, rescues, floods, and pre-hospital care. The data is classified into four primary event types: pre-hospital care, fires, search and rescue, and hazardous material incidents, reflecting the multifaceted nature of emergency responses in this region.

To evaluate the generalizability of our approach, we also tested it in two distinct macro-regions: Macro Cha (Chapecó region) and Macro Blu (Blumenau region). These regions differ significantly in weather patterns, types and frequency of occurrences, cultural behavior, and other contextual factors. Macro Cha includes the cities of Chapecó, Xaxim, and Xanxerê, while Macro Blu comprises Blumenau, Gaspar, and Timbó.

Figure 30 (top left) illustrates the geographical distribution of these macro-regions. The yellow marker represents Macro Cha, the green marker indicates Macro Blu, and the blue marker denotes Macro Flo. By incorporating these diverse regions, we assess the robustness and adaptability of our model to region-specific emergency patterns (VIEIRA ROQUE et al., 2025). Although traffic conditions are a potentially valuable feature for emergency forecasting, particularly in urban areas where mobility affects response times, we did not incorporate traffic data in this study. The primary

reason is the absence of publicly available traffic datasets that align in both temporal resolution and geographic coverage with emergency occurrence data. Multiple attempts were made to identify suitable sources, but none provided consistent or historical traffic data that could be integrated with our multi-year, multi-region dataset.

5.2.1.1 Real MFC Occurrences

The data, obtained from the MFC (Military) system, includes details about each event, such as location, type, and date. Table 15 summarizes sample entries categorized by city and event type. During preprocessing, duplicate entries were removed, and missing location data were imputed using neighborhood-based values to maintain data integrity. Macro-Flo stands out with the highest numbers compared to other regions, particularly in pre-hospital care occurrences, where Florianópolis leads due to being the largest city in the state. Notably, some cities, such as Timbó and Biguaçu, have a relatively low number of recorded events. However, despite their minor counts, including these cities in their respective macro regions is essential, as they are interconnected within a broader network of neighboring cities.

Table 15 – Number of occurrences for each city across the studied macro-regions.

Region	Fire	Inspection	Pre-hospital	Search & Rescue
Biguaçu	388	317	3417	206
Florianópolis	3138	2126	34702	3063
Palhoça	1155	743	11506	590
São José	988	715	10636	614
Chapecó	4652	199	19894	855
Xaxim	218	181	7173	246
Xanxerê	562	279	9098	597
Blumenau	853	296	21038	6066
Gaspar	252	185	4198	1038
Timbó	120	181	4925	725

Figure 31 illustrates the daily distribution of occurrences in Macro-Flo, highlighting temporal variations throughout the analyzed period. Pre-hospital care incidents are the most frequent across all cities, as they represent the primary type of emergency attended by the MFC. A noticeable increase in occurrences is observed during the summer months (December–March), likely due to these cities' locations along the western coast near the sea, making them popular vacation destinations.

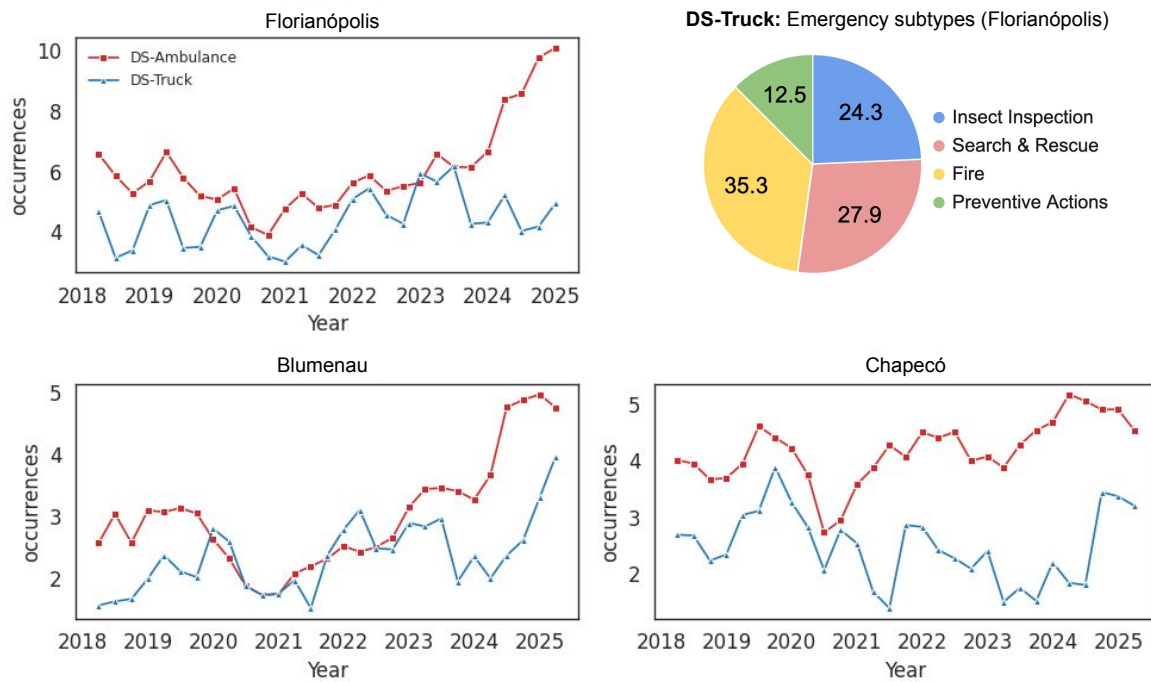


Figure 31 – Daily occurrences for Macro-Flo.

5.2.1.2 Weather Data

Relative Humidity (rhum): Represents the moisture percentage in the air relative to the maximum amount the air can hold at a given temperature. High values indicate a moist atmosphere, while low values correspond to drier conditions. **Wind Direction (wdir):** Indicates the direction from which the wind originates, measured in degrees from true north. This variable provides insight into the prevailing wind patterns and potential weather fronts. **Precipitation (prcp):** Measures the amount of liquid or frozen water that falls to the ground, recorded in millimeters. It encompasses rainfall, snow, sleet, and hail. **Pressure (pres):** Atmospheric pressure at sea level, recorded in hectopascals (hPa). Changes in pressure are associated with different weather conditions, such as high-pressure systems leading to clear skies and low-pressure systems indicating storms. **Wind Speed (wspd):** Denotes the average speed of the wind, measured in meters per second (m/s). This variable is important to understanding the weather dynamics and the storms' intensities. **Temperature (temp):** The ambient air temperature, recorded in degrees Celsius (°C). It is one of the most fundamental weather parameters influencing daily weather conditions. **Dew Point (dwpt):** The temperature at which air becomes saturated with moisture, leading to the formation of dew. It is an indicator of atmospheric moisture content.

These variables, depicted in Figure 32, were chosen for their relevance to weather dynamics and their potential impact on emergency events. To integrate weather data with emergency occurrences, we used each city's location to retrieve weather conditions and fused this data with the occurrences based on their corresponding date. The `prcp` variable, which represents precipitation levels, was incomplete in the dataset, with the first two years' values and instances recorded as zero. To ensure a more accurate representation of precipitation, we applied a regression-based imputation technique.

In this approach, we trained a linear regression model using available weather features, such as `temp`, `rhum`, `wspd`, `pres`, `wdir`, `dwpt`, and cyclical calendar features. The model was trained on instances where `prcp` was neither missing nor zero, learning patterns from valid precipitation data. Then it was used to predict and replace missing and zero values, ensuring that precipitation estimates align with the overall weather conditions.

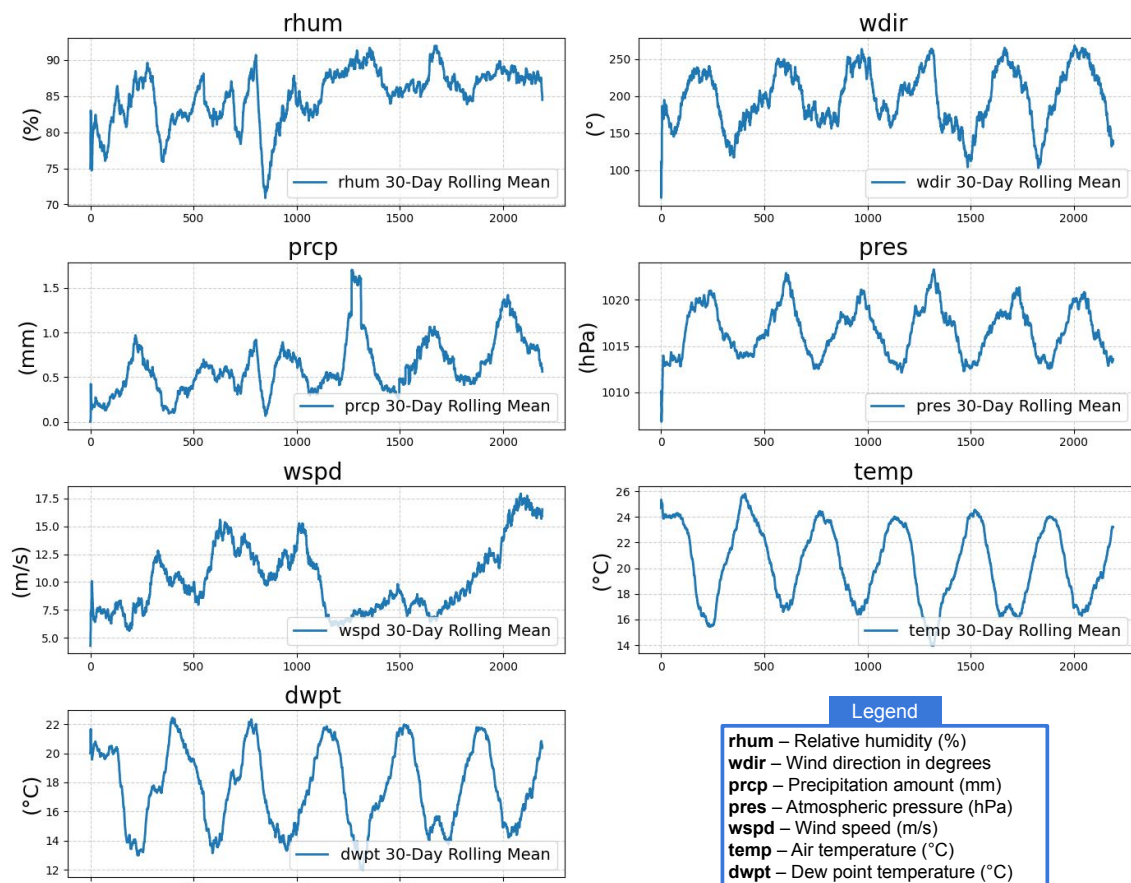


Figure 32 – Weather variables.

5.2.1.3 Calendar Days Data

The special days data includes features that represent both holidays and temporal information, such as the day of the week and the month. This information was fused with the dataset using the date as the key, ensuring alignment with the weather data and emergency occurrences. The variables are described below:

Holidays: Specific public holidays and significant local events were identified for inclusion in the dataset. These days often influence human activity patterns and emergency occurrence rates. **Day of the Week:** Each day was encoded as a categorical variable to distinguish between weekdays and weekends. Additionally, sinusoidal transformations (`day_of_week_sin` and `day_of_week_cos`) were applied to capture cyclic patterns within a week. **Week of the Year:** To account for seasonal periodicity, we included `week_sin` and `week_cos`, which provide continuous representations of the week's position within the year. **Weekend Indicator:** A binary variable (`weekend`) was introduced to differentiate between weekends and weekdays, considering the impact of non-working days on emergency occurrences. **Month:** The dataset includes the month of the year, capturing seasonal trends that may influence weather conditions and emergency patterns.

Figure 33 illustrates the distribution of holidays in 2019 and their corresponding emergency occurrences, emphasizing the relevance of temporal and special day features in the analysis. The year 2019 was selected as it serves as a representative period within the dataset. The figure reveals that certain holidays are associated with an increase in emergency incidents, for example, the New Year, Carnival, Easter, and Christmas, all in blue. These holidays show a clear upward trend in the number of occurrences, which may be attributed to increased population movement during travel or large-scale festive gatherings commonly held in the region during these times.

5.2.2 Regionalization

As in Experiment 1, we separate the dataset into two operational categories, ambulance-related (DS-Ambulance) events and truck-related (DS-Truck) events. Spatial segmentation is critical for this problem, as each area can have specific behavior. For each of the datasets (DS-Ambulance and DS-Truck), we perform the segmentation; in this way, we can separate the two behaviors of the types of occurrence. Figure 34 illustrates how we employ the step of the framework for this problem with two distinct datasets.

For this work, we evaluated 20 different values of groups to define what best

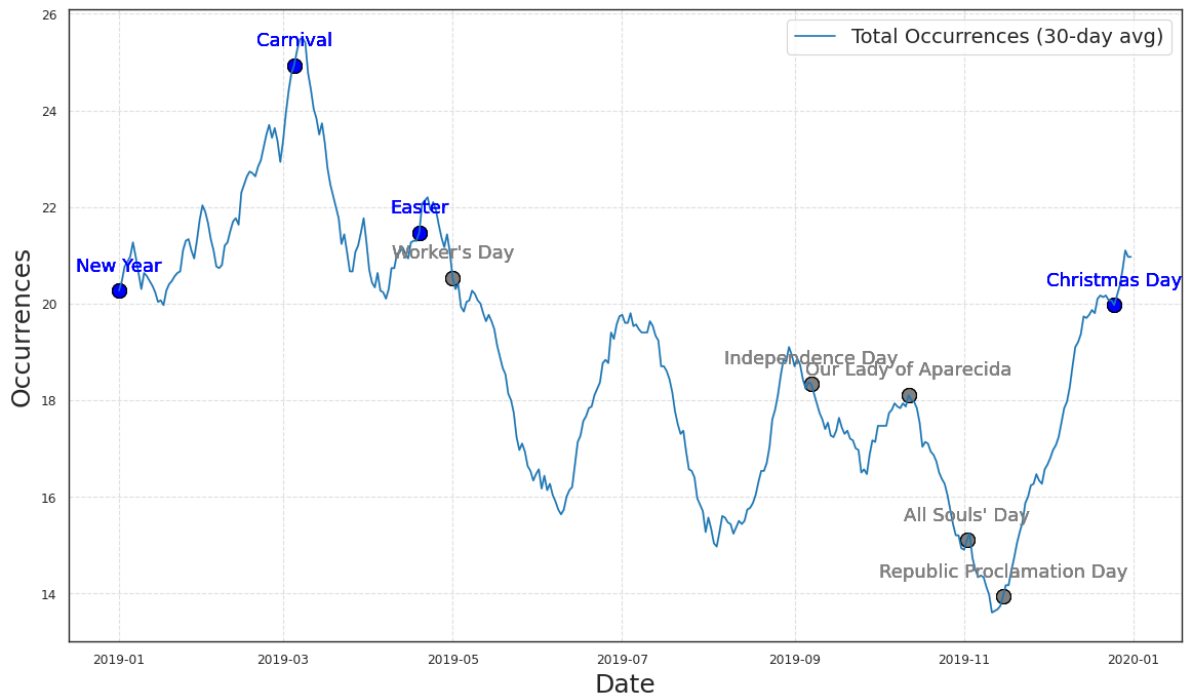


Figure 33 – Daily Total Occurrences (2019) in Macro-Flo with Special Days.

separates the observations in the problem. We defined the number 4 as the smallest possible value for this analysis because the fire department has four emergency dispatchers spread throughout the Florianópolis macroregion. The idea of testing higher values is that this could indicate the need to build more stations if this improves the analysis and the efficiency of rescue operations. In our Silhouettes' evaluation, the best value of groups was 8 for k-means and k-Medoids for DS-Ambulance and 7 for DS-Truck occurrences. For Agglomerative the best value was 5. The results indicate that dividing the city into eight and seven regions leads to a more homogeneous allocation.

With the definition of k for the segmentation of locations, the initial series was divided into eight series for DS-Ambulance and seven for DS-Truck events, defined by location (see in Figure 35). The figure demonstrates that k-means effectively segments observations based on geographical coordinates (Longitude and Latitude). The region has a different population distribution throughout its area. There are denser regions, like where we can see the green, blue, and purple dots. The other regions are less populated; however, they also have urbanized areas amidst native forests. For the DS-Truck, the figure representing the segmentation by k-means was very similar to Figure 35. In this way, we prefer not to insert it in this work.

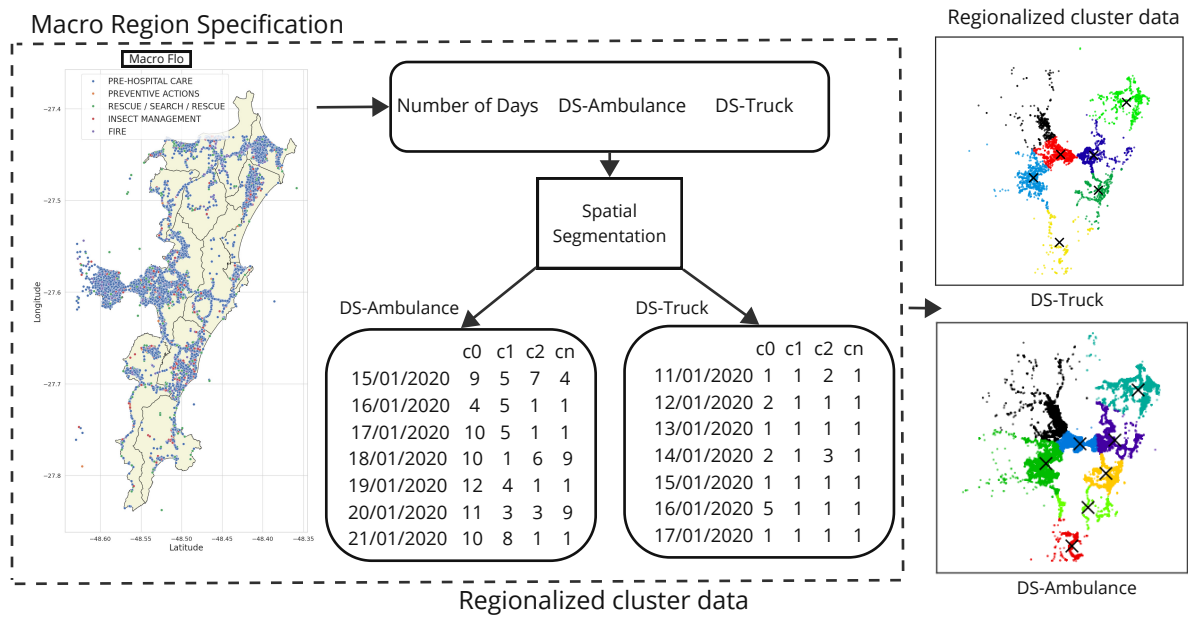


Figure 34 – Spatial segmentation procedure to generate regions data for the models.

We can observe that c2, c3, and c4 in Figure 35 have more observations over time. The c0, c1, and c6 have very similar behavior to each other. Clusters c5 and c7 have fewer observations. This can be explained by the team's service region (the location of the rescue team), where c2, c3, and c4 are regions with a greater concentration of people than the other regions. On the other hand, c6, c4, c2, and c1 have the bigger number of observations. This indicates that occurrences of this type are more frequent than those observed in the DS-Ambulance for this region.

5.2.3 TFT for emergency prediction

This section presents the results obtained using our custom-designed, optimized TFT architecture. We also compare them with classic time series prediction and machine learning techniques. The experiments were conducted on a system running Ubuntu 22.04, equipped with an Intel 13th Generation Core i5-13420H processor featuring eight cores and a clock speed of up to 4.6 GHz. The system also included an NVIDIA GeForce RTX 3050 GPU with 6GB of dedicated memory, 16GB of DDR5-5600 MHz RAM, and a 512GB Solid State Drive (SSD) for storage.

5.2.3.1 TFT for Predicting Emergency Events

The TFT model was trained over 350 epochs, with performance assessed using the Least Absolute Deviations (L1) loss metric. The Adam optimizer managed

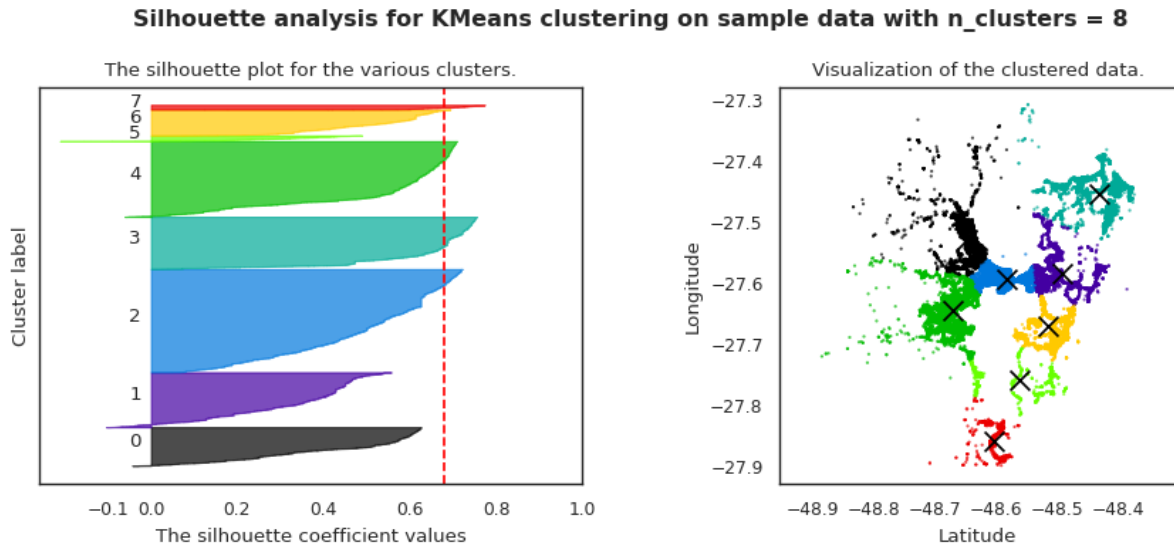


Figure 35 – Best k value for the problem and its visual separation seems from silhouettes for DS-Ambulance.

the learning rate and minimized the average loss. Time Series cross-validation with five folds was implemented to ensure robustness and prevent overfitting. The dataset was split into 70% for training, 20% for validation, and 10% for testing. The experiment was fixed with an encoder length (e_l) of 7 days and a prediction length (p_l) of 1 day. For hyperparameter optimization, we utilized Optuna, a powerful tool that employs the TPE to minimize the loss function. This approach was particularly beneficial for optimizing the numerous hyperparameters of the TFT. The optimization phase prioritized some hyperparameters to enhance model performance, including the h_s , l_r , d_r , number of attention heads (n_{ah}), and hidden continuous size (h_{cs}).

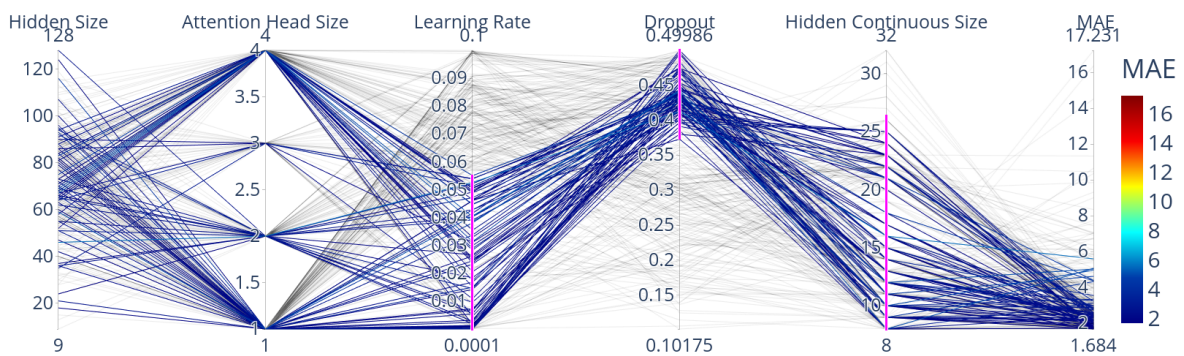


Figure 36 – Optimization results for DS-Ambulance occurrences.

We conducted 350 trials for each dataset and each experimental approach

to identify the optimal parameter configurations. Figure 36 provides a visual representation of the optimization process for the DS-Ambulance dataset. The lr tends towards lower values, with the optimization process converging to 0.0001 in the final stage. Higher dr yield improved results, indicating their effectiveness in regularizing the model. The hcs appears to have a minimal impact on the MAE error, suggesting it is less critical for performance. The figure also highlights the significance of hyperparameter tuning, as suboptimal configurations can lead to errors more than twice as high as the optimized settings. The best parameter for the configuration of the TFT model can be found in 12.

Table 16 – Parameters optimization for different approaches.

Model	Parameters	DS-Ambulance	DS-Truck
ARIMA	p, q, d	$p=6, q=1, d=1$	$p=5, q=1, d=1$
MA	SO	7	7
DT	md	4	3
MLP	af, hls	tanh, 100	tanh, 50
SVM	c, ϵ	$c=0.1, \epsilon=0.1$	$c=0.1, \epsilon=0.2$
LSTM	lr, dr, hs, nhl	0.0016, 0.45, 19, 1	0.0012, 0.47, 9, 3
TFT	hs, lr, dr, nah, hcs	22, 0.00013, 0.41, 1, 11	26, 0.00015, 0.13, 2, 20
R1-LSTM	Fixed	Fixed	Fixed
R2-LSTM	Fixed	Fixed	Fixed

5.2.3.2 Performance Comparison of Models

The performance of two time series (TS) forecasting techniques MA and ARIMA, and three traditional machine learning techniques: SVM (CORTES; VAPNIK, 1995), decision trees (DT) (SAFAVIAN; LANDGREBE, 1991), and MLP (PAL; MITRA, 1992), was compared with that of the TFT proposed in this research. The experiments for TS analysis were conducted using the statsmodels (SEABOLD; PERKTOLD, 2010) module, while those for machine learning utilized the scikit-learn (BUTINCK et al., 2013) package in Python. We also compare TFT performance with an LSTM approach. All TS and machine learning techniques were evaluated using 5-fold cross-validation to ensure a fair comparison with the TFT models. These models were configured with the same experimental setup as the proposed approach, using an $e/$ of 7 days and a $p/$ of 1 day. This ensures that each model learns and predicts under equivalent condi-

tions, enabling a direct performance comparison. Table 18 summarizes the results for different models.

LSTM model optimization focused on the following hyperparameters: hs , lr , dr , and nhl . The experimental designs for TS models were first established. An analysis of the autocorrelation function was performed to determine the point where it approaches zero, thereby selecting the optimal smoothing order for MA, set to 7, as it approximated the average. The parameters for the ARIMA method were optimized using the `autoarima` (SMITH et al., 2017) package. For machine learning techniques, an exhaustive grid search was used to find the optimal parameters. We can also see the best parameter configurations for all models compared in 16.

The comparison of different predictive models for emergency event prediction in the DS-Ambulance and DS-Truck datasets highlights the effectiveness of the TFT model. Table 17 presents the MAE, MSE, and RMSE for each approach, along with their respective prediction times. The results indicate that the TFT model consistently outperforms traditional time series and machine learning models in predictive accuracy, particularly for the DS-Truck dataset.

For the DS-Ambulance dataset, the TFT model achieves an MAE of 1.58, highlighted in blue. These results suggest that the TFT model effectively captures temporal dependencies while maintaining superior generalization capabilities. In the DS-Truck dataset, the superiority of the TFT model is even more pronounced. It achieves the lowest MAE (0.40), significantly outperforming, demonstrating its robustness in minimizing forecasting errors. However, the computational cost of the TFT model for DS-Truck is higher (4.5 ms) compared to most other models, including LSTM. Still in Table 17, we compare our model with a competing solution from the literature. We train an LSTM with the same parameters from a reference that also predicts daily events (CORTEZ et al., 2018) (named R1-LSTM on the table) and (VIEIRA ROQUE et al., 2025) (named R2-LSTM on the table). As the results show, the TFT proposed in our work is superior to this approach for both datasets (DS-Ambulance and DS-Truck) in terms of error.

The results confirm that while classical models such as ARIMA struggle with non-linear patterns and machine learning models such as DT and SVM achieve moderate success, deep learning models, particularly TFT offer the most accurate and reliable predictions. The TFT model's ability to leverage attention mechanisms and multi-horizon prediction capabilities enables it to capture complex temporal relationships effectively. These findings reinforce the suitability of TFT for predictive tasks in predicting emergency events, where high accuracy is crucial to effective decision-

making.

Table 17 – Results for different approaches.

Method	DS-Ambulance			DS-Truck			
	MAE	MSE	RMSE	MAE	MSE	RMSE	T
LSTM	1.67	6.20	2.49	0.74	0.78	0.89	2.9
TFT	1.58	5.04	2.24	0.40	0.38	0.61	4.5
ARIMA	2.86	9.00	3.00	0.64	0.60	0.78	9.14
DT	2.06	7.06	2.54	0.55	0.46	0.68	1.30
MLP	2.56	10.90	3.30	0.52	0.49	0.70	0.65
SVM	2.17	7.65	2.65	0.46	0.43	0.63	0.58
R1-LSTM	2.08	6.35	2.52	1.01	1.67	1.29	2.63
R2-LSTM	2.28	7.65	2.77	1.00	1.64	1.28	1.52

5.2.3.3 Multistep prediction with TFT model for different $e/$

We compared multistep predictions using the TFT and LSTM models across different $e/$ and p/s . The configurations chosen for evaluation included combinations of $e/$ of 7, 21, and 30 days, and prediction horizons of 1, 5, and 7 days. This setup allowed us to assess the impact of historical input length on prediction error for two datasets: DS-Ambulance and DS-Truck.

Overall, the results presented in Table 18 demonstrate that the TFT model consistently outperforms LSTM in most scenarios, particularly for shorter prediction horizons. Short-Term Encoders (7 Days): TFT achieves lower MAE and MSE across different p/s than LSTM. In particular, for the DS-Truck dataset, the MAE for TFT in the (7,1) setup is 0.40 compared to 0.74 for LSTM, indicating superior short-term predictive capacity. Similar trends are observed for 5-day and 7-day forecasts. Medium-Term Encoders (21 Days): While both models improve in accuracy with increased $e/$, TFT maintains a performance advantage. For example, in the (21,5) configuration, TFT records an MAE of 1.60 and 0.59 for DS-Ambulance and DS-Truck, respectively, compared to 1.62 and 0.67 for LSTM. Long-Term Encoders (30 Days): The results suggest that extending the $e/$ to 30 days leads to further improvements, particularly in DS-Ambulance. TFT shows a significant reduction in MAE and MSE when compared to LSTM, with the most notable improvement seen in the (30,1) setup, where TFT achieves an MAE of 0.78 versus LSTM's 1.27. However, for DS-Truck, the differences are less pronounced in longer horizons (5 and 7 days), suggesting diminishing returns with increasing $e/$.

These findings highlight the robustness of the TFT model in handling multistep predictions in different el . The ability of TFT to capture long-term dependencies while effectively weighting relevant features contributes to its superior performance over LSTM. Furthermore, the results suggest that while increasing el generally improves MAE error, the gains diminish beyond a certain point, particularly in datasets with lower variability such as DS-Truck. This capability is particularly relevant in emergency scenarios, where anticipating demand several days in advance is crucial for allocating personnel and vehicles, pre-positioning resources, and improving decision-making efficiency. The TFT consistently outperformed LSTM in the short, medium, and long term, especially with el of 7 to 30 days and prediction windows of up to 7 days. This suggests that TFT is well-suited for operational planning, where both near-term response and medium-range forecasting are necessary.

Table 18 – Results for different el and pl windows in DS-Ambulance and DS-Truck datasets.

		DS-Ambulance			DS-Truck		
$el-pl$		MAE	MSE	RMSE	MAE	MSE	RMSE
7-1	LSTM	1.67	6.20	2.49	0.74	0.78	0.89
	TFT	1.58	5.04	2.24	0.40	0.38	0.61
7-5	LSTM	1.67	4.92	2.22	0.67	1.14	1.07
	TFT	1.57	4.56	2.14	0.66	1.06	1.03
7-7	LSTM	1.41	3.35	1.83	0.68	1.00	1.00
	TFT	1.41	3.29	1.81	0.65	1.03	1.02
21-1	LSTM	1.61	5.78	2.40	0.66	0.73	0.85
	TFT	1.60	5.65	2.38	0.59	0.71	0.84
21-5	LSTM	1.62	4.74	2.18	0.67	1.04	1.02
	TFT	1.60	4.53	2.13	0.59	0.48	0.69
21-7	LSTM	1.47	3.41	1.85	0.69	0.99	1.00
	TFT	1.50	3.31	1.82	0.67	1.06	1.03
30-1	LSTM	1.27	2.24	1.50	0.68	0.74	0.86
	TFT	0.78	1.31	1.14	0.59	0.70	0.84
30-5	LSTM	1.57	4.51	2.12	0.67	1.09	1.04
	TFT	1.58	4.58	2.14	0.62	1.11	1.06
30-7	LSTM	1.49	3.41	1.84	0.68	1.06	1.03
	TFT	1.45	3.15	1.77	0.67	1.09	1.04

5.2.3.4 TFT in different Macro regions

The evaluation of the TFT model was extended to two distinct macro-regions, CHA and BLU, in addition to the FLO dataset. The entire pipeline was executed for

Table 19 – MAE for different macro-region scenarios using DS-Ambulance and DS-Truck datasets for 1 day *pl.*

Macro-Region	MAE (DS-Ambulance)	MAE (DS-Truck)
Flo	1.58	0.40
Cha	1.99	0.89
Blu	1.12	0.44

both new datasets, CHA and BLU, to assess the generalizability of the methodology. The number of optimal clusters was determined to be four for both CHA and BLU. Table 19 presents the MAE for each macro-region using DS-Ambulance and DS-Truck datasets with a one-day prediction horizon. The results highlight the regional differences in predictive performance, demonstrating variations in MAE in different geographic areas. The FLO region, which was used as the initial dataset, achieved an MAE of 1.58 for DS-Ambulance and 0.40 for DS-Truck. When applied to CHA, the model showed a higher MAE of 1.99 for DS-Ambulance and 0.89 for DS-Truck, indicating a potential increase in the difficulty of prediction in this region. In contrast, the BLU region exhibited improved predictive performance with an MAE of 1.12 for DS-Ambulance and 0.44 for DS-Truck.

These findings suggest that the model's effectiveness varies across macro-regions, likely influenced by region-specific factors such as traffic patterns, incident reporting consistency, and environmental conditions. The results emphasize the importance of tailoring predictive models to account for regional differences to improve the prediction error. Additionally, the optimal clustering of four clusters in both CHA and BLU suggests a stable segmentation approach, reinforcing the robustness of the methodology across different geographic contexts. This paves the way for scalable adoption of our pipeline across municipalities, with minor retraining or adaptation, thereby enhancing emergency responsiveness at a broader geographic scale.

5.2.4 Interpretability and Evaluation

5.2.4.1 Ablation analysis

The ablation study presented in Table 20 evaluates the impact of different data components on the model's predictive performance. In DS-Ambulance, removing weather data while retaining occurrences and calendar features results in a slight increase in MAE from 1.58 to 1.65, while excluding holiday information yields an MAE of 1.66. The highest error occurs when only occurrences are used (1.70 MAE),

suggesting that external factors significantly enhance predictive performance. A similar trend is observed in DS-Truck, where using all features provides the best MAE (0.40). Removing weather data leads to a notable increase in MAE to 0.64, while excluding calendar or holiday data results in even higher errors (0.63 and 0.78, respectively). The worst performance occurs when only occurrences are used, with an MAE of 0.89, reinforcing the need for contextual data. Interestingly, using only weather data produces an even higher MAE (0.92), indicating that while weather alone has limited predictive power, its combination with occurrences provides valuable improvements.

In general, these findings highlight the importance of integrating multiple data sources in predicting emergency events. The model's predictive capability declines when key external factors such as calendar effects, holiday influence, and weather conditions are removed. This underscores the advantage of using a holistic approach in time series prediction, where contextual variables enhance the model's ability to capture temporal dependencies and improve accuracy.

Table 20 – Removing data from the TFT model to test predictability.

Combinations	DS-Ambulance			DS-Truck		
	MAE	MSE	RMSE	MAE	MSE	RMSE
all	1.58	5.04	2.24	0.40	0.38	0.61
occur + weather	1.68	4.68	2.16	0.64	0.96	0.98
occur + calendar	1.65	5.08	2.25	0.63	0.91	0.95
occur + holiday	1.66	4.91	2.21	0.78	0.81	0.90
occurrences	1.70	4.99	2.23	0.89	1.04	1.01
weather	1.68	5.77	2.40	0.92	1.15	1.07
calendar	1.63	5.23	2.28	0.88	1.05	1.02

5.2.4.2 Interpretability for different scenarios

We analyze interpretability outputs under different scenarios, varying the e_l . We compare scenarios with shorter (7 days) and longer (30 days) historical input windows to investigate how attention and variable importance change depending on the amount of past information available. TFT calculates variable importance by leveraging its Variable Selection Networks to assign dynamic attention-based weights to each feature and aggregates attention scores across time steps and heads to quantify their overall contribution to the prediction (LIM et al., 2021).

Figure 37 presents the attention weights generated by the model for two different e_l : Figs 37a 7 days and 37b 30 days. These visualizations help us to understand

which time steps in the input sequence the model considers most relevant when forecasting emergency occurrences. In Figure 37a, we observe that the attention values increase as the input approaches the prediction day, indicating that the model places greater importance on more recent observations. A similar pattern is evident in Figure 37b, where attention gradually intensifies over time, again highlighting the model's tendency to focus more on recent data points when making predictions.

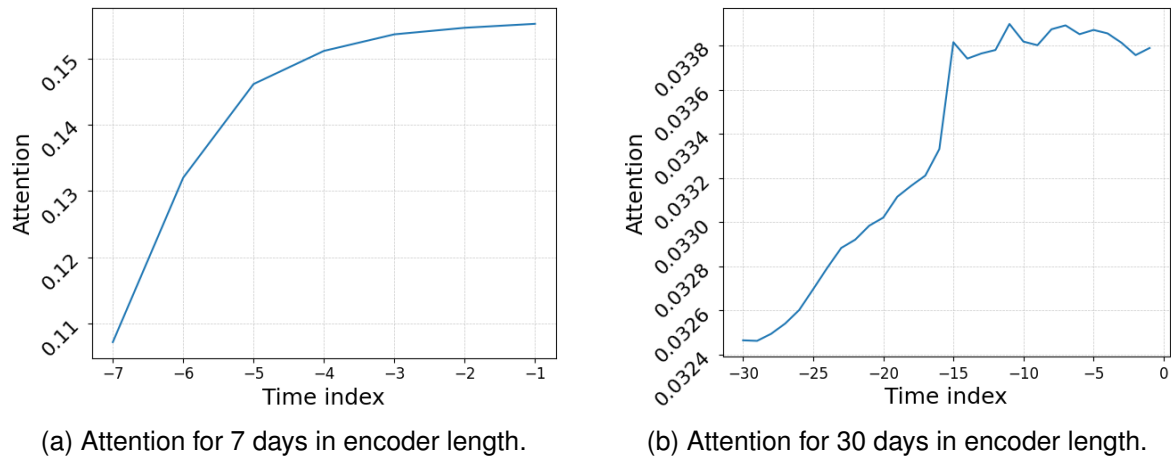
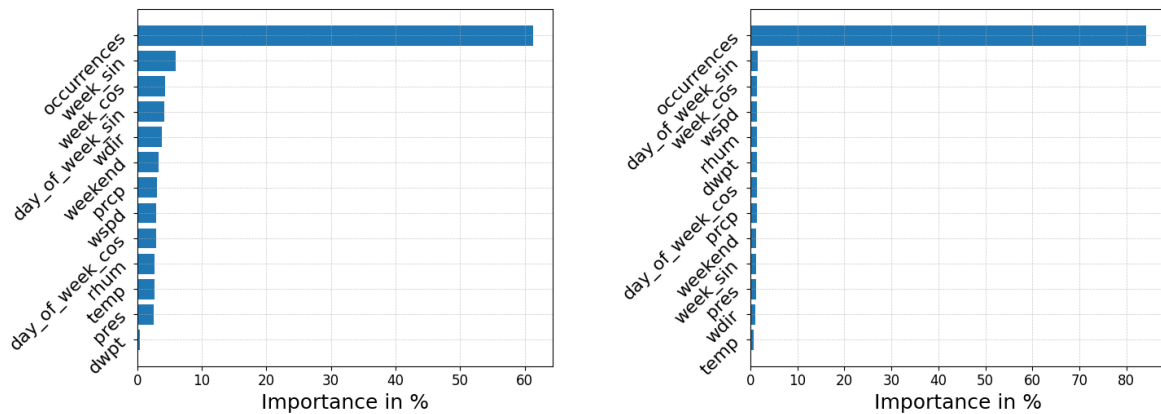


Figure 37 – Attention in different scenarios.

Figure 38 shows the variable importance derived from the encoder for two scenarios. In the shorter encoder window (7 days), Figure 38a, the most influential variables are weekend (indicating whether the day falls on a weekend) and wdir (wind direction), each contributing approximately 14% to the model's prediction process. Interestingly, when the e_l is extended to 30 days (Figure 38b), the most important variable shifts significantly to occurrences (historical emergency counts), which accounts for around 50% of the overall importance. This suggests that, with a longer historical window, the model learns to rely more heavily on temporal patterns within the target variable itself. In contrast, occurrences are among the least important features in the 7-day encoder setup, highlighting how the model's feature dependency changes with the amount of contextual information provided.



(a) Encoder length for 7 days in encoder length. (b) Encoder length for 30 days in encoder length.

Figure 38 – Encoder length in different scenarios.

Finally, Figure 39 presents the variable importance for the decoder stage of the model. In the 7-day scenario (Figure 39a), weather-related variables dominate the top positions in terms of importance, indicating that short-term weather conditions play a key role in guiding the model's predictions. In contrast, in the 30-day scenario (Figure 39b), calendar-related features, such as day of the week and month, take on greater importance, except for *dwpt* (dew point temperature), which holds the highest importance overall. This shift suggests that with a longer historical context, the model relies more on seasonal and temporal patterns than on immediate weather conditions. From a decision support point of view, the interpretability features of the TFT provide transparency in understanding which features contribute most to predictions. This is invaluable for stakeholders in emergency response services, enabling data-driven trust and insightful planning based on model outputs.

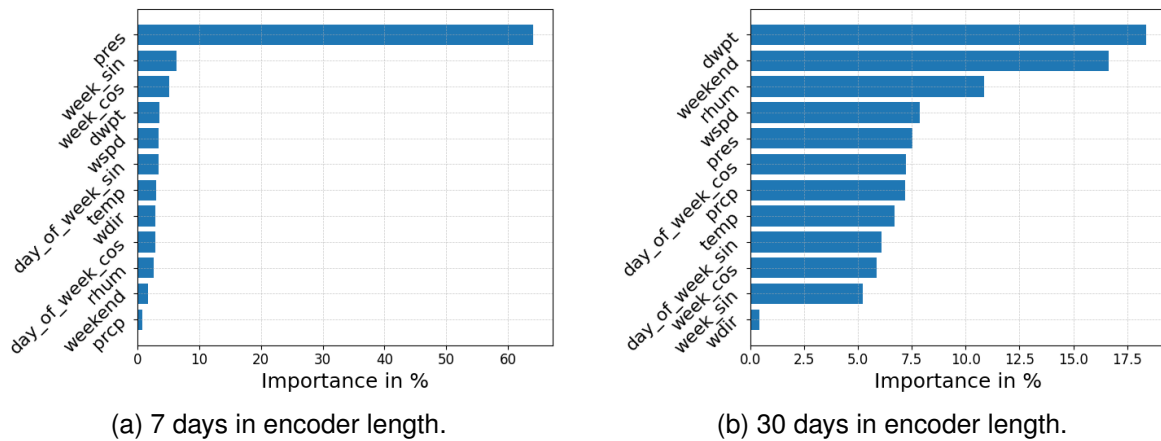


Figure 39 – Decoder length in different scenarios.

5.3 CASE STUDY: REPRODUCING AN EMERGENCY SCENARIO WITH TFT MODEL

This case study demonstrates the application of a TFT model to reproduce and anticipate emergency scenarios using historical occurrences and multimodal data for Macro-Flo Region. By predicting occurrences for different clusters, the analysis provides operators with actionable insights to support proactive resource allocation and risk mitigation strategies.

Figure 40 presents the predicted number of occurrences for each cluster up to 15 days ahead. Each cluster represents a spatial-temporal grouping of events, and the predictions are color-coded according to the expected magnitude of occurrences. This visualization allows operators to identify which clusters are likely to experience higher activity, enabling targeted interventions, optimized deployment of emergency teams, and improved preparedness for high-risk periods. We also present the Table 21 that present the error in terms of MAE, MSE, and RMSE in this scenario of 15 days.

Figure 41 maps the past 60 days of occurrences by cluster, displaying their geographic distribution. Using the same color scale derived from the predicted cluster values, this map highlights areas with higher risks, which correlate with clusters expected to see elevated occurrences in the coming days. By overlaying prediction-based cluster information on the geographic map, operators can quickly identify hazardous regions, anticipate the demand for emergency services, and prioritize monitoring or preventive actions.

Together, these two figures provide complementary perspectives: the first quantifies future occurrences temporally, while the second contextualizes risk spatially. This

combination enhances situational awareness, supports data-driven decision-making, and allows operators to plan more effectively, reducing response times and improving the overall efficiency of emergency management.

Table 21 – Error Metrics related to actual predictions to different clusters .

cluster	MAE	MSE	RMSE
0	2.42	8.38	2.89
1	1.68	4.92	2.22
2	2.34	7.45	2.73
3	1.48	4.31	2.08
4	0.58	0.73	0.85
5	1.98	6.39	2.53
6	1.72	3.92	1.98
7	1.20	2.06	1.43

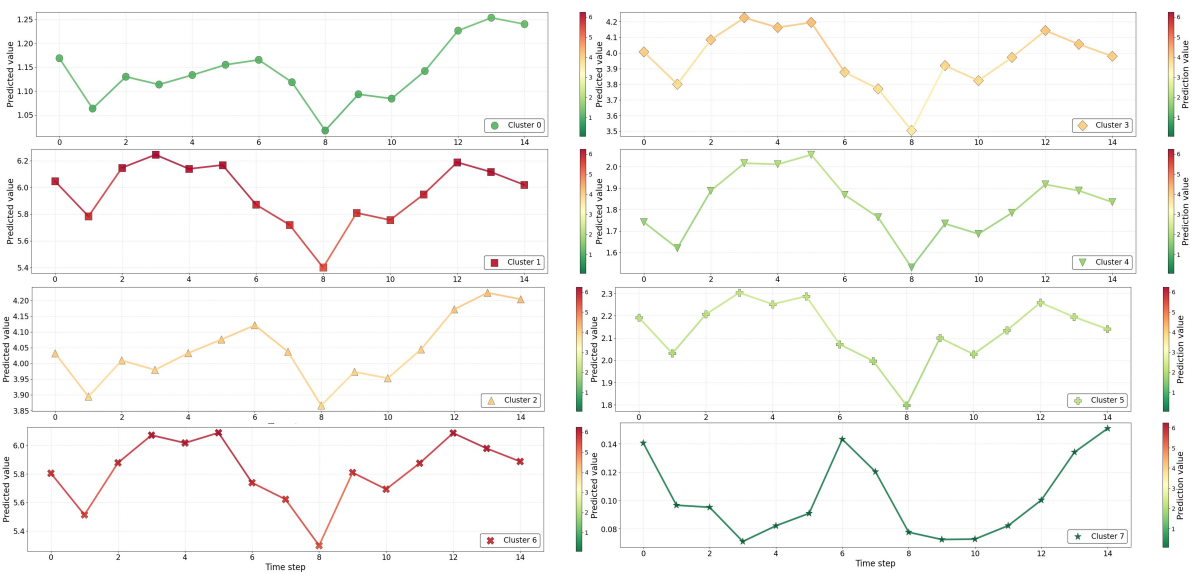


Figure 40 – Predictions by cluster 15 days ahead to help decision making.

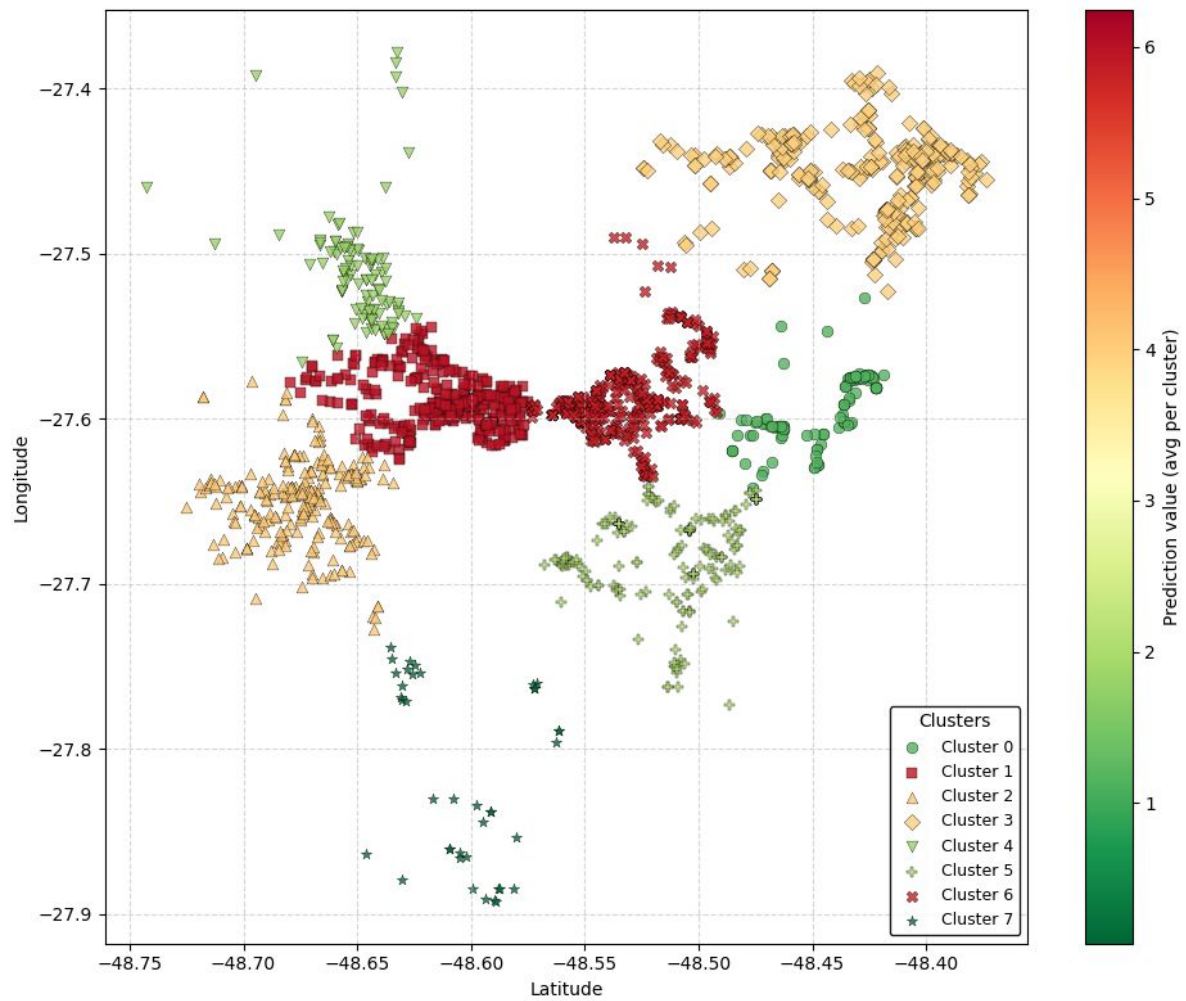


Figure 41 – Past 60 days showing the regions that are dangerous to occurrences in the next days.

6 DISCUSSION

This chapter examines the principal findings of the thesis, linking the discovered results with existing literature.

6.1 A FRAMEWORK PROPOSAL FOR EMERGENCY EVENTS PREDICTION

The first case study, addressed in Vieira Roque et al. (2025), in Florianópolis demonstrated the effectiveness of using an LSTM model for predicting daily emergency events. The results, as shown in Table 11, clearly indicate that the optimized LSTM model significantly outperformed traditional time series models (ARIMA, STL, MA) and other machine learning techniques (DT, MLP, SVM) across both DS-Ambulance (pre-hospital care) and DS-Truck (other emergencies) datasets, achieving the lowest MAE, MSE, and RMSE. This highlights the capability of recurrent neural networks in capturing the temporal dependencies inherent in emergency event data. We also confirm that regionalized models outperform generalized ones, as demonstrated in Section 5.1.4.4, where localized training consistently yields lower prediction errors.

The evaluation of multi-step forecasting with the LSTM model (Table 13) indicated that predicting one day ahead yielded the best performance, which aligns with the optimization strategy focused on a look-ahead of one day. The comparison between the recursive and direct prediction methods showed the recursive approach generally performing better for multiple look-ahead windows, suggesting that feeding the model's output back as input for subsequent predictions is a viable strategy for short-term forecasting.

The second case study, employing a TFT model for macro-regional experiments, further validated the effectiveness of the FORESEE framework across different geographical areas (Macro Flo, Macro Cha, and Macro Blu). The TFT model consistently outperformed traditional time series and machine learning models, including LSTM, as evidenced in Table 17. Notably, for the DS-Truck dataset in the Macro Flo region, the TFT model achieved a significantly lower MAE compared to all other methods.

The multi-step prediction analysis with the TFT model (Table 18) demonstrated its robustness across different encoder lengths (7, 21, and 30 days) and prediction horizons (1, 5, and 7 days). The TFT model generally outperformed LSTM, particularly for shorter prediction horizons, indicating its strength in capturing both short-term and medium-term dependencies in the data.

The ablation study (Table 20) provided compelling evidence for the importance of integrating multimodal data sources. The results clearly showed that the predictive performance of the TFT model declined significantly when weather, calendar, or holiday data were excluded, emphasizing the value of data fusion in uncovering critical correlations and patterns. The most substantial performance drop occurred when only historical emergency event data was used, underscoring the necessity of incorporating external contextual information.

The interpretability analysis of the TFT model, through attention weights and variable importance (Figures 37, 38, and 39), offered valuable insights into the factors influencing emergency event predictions. The model dynamically adjusted its attention based on the encoder length, focusing more on recent data for shorter windows and relying more on historical occurrences for longer windows. The variable importance analysis highlighted the significance of factors like weekends and wind direction in shorter encoder settings, while historical occurrences and calendar features became more influential with longer historical contexts. Weather-related variables played a crucial role in the decoder stage for short-term predictions, while calendar features gained importance for longer horizons.

The application of the TFT model to different macro-regions (Table 19) revealed variations in predictive performance, suggesting the influence of region-specific characteristics on emergency event patterns. While the Macro Flo region showed strong performance, the model exhibited different levels of accuracy in Macro Cha and Macro Blu, highlighting the need for region-specific considerations in emergency forecasting models, as proposed by the FORESEE framework.

6.2 DISCUSSION OF THE FORESEE FRAMEWORK

The results from both case studies provide strong support for the key components of the FORESEE framework.

- **Regionalized modeling using spatially-segmented data:** The segmentation of Florianópolis and the evaluation across different macro-regions demonstrated that considering the spatial context and building models for specific areas can lead to more accurate predictions compared to a single generalized model. The improved performance of the LSTM model on spatially segmented data in the first case study and the varying performance of the TFT model across different macro-regions in the second case study underscore the importance of this

aspect of the framework.

- **Effective handling of multimodal data types through data fusion:** The significant improvement in prediction accuracy observed with the TFT model when integrating weather and calendar data, as shown in the ablation study, validates the effectiveness of the data fusion component of the FORESEE framework. Combining diverse data sources provided a more comprehensive understanding of the factors influencing emergency events.
- **Multi-step forecasting:** Both the LSTM and TFT models demonstrated the capability of multi-step forecasting, providing predictions for future time windows. The analysis of different prediction horizons, particularly with the TFT model, showed the framework's potential to offer insights for short, medium, and potentially longer-term planning in emergency management.

6.3 ADDRESSING THE RESEARCH QUESTION

RQ: Can machine learning techniques and statistical methods predict emergency events across different temporal horizons, thereby improving decision-making when combined with multimodal data sources?

The evidence from the two case studies strongly suggests that machine learning and statistical techniques can indeed accurately predict emergency events in different horizons, especially when integrated with multimodal data sources and applied within a regionalized framework like FORESEE. The LSTM model showed promising results for short-term predictions in a specific city, while the TFT model demonstrated superior performance for multi-step forecasting across a macro-region when incorporating meteorological and calendar-based features.

The results from the case study presented in Section 5.3 illustrate the practical impact of such models on operational decision-making and resource allocation. The TFT model, applied to the Macro-Flo region, successfully reproduced historical emergency dynamics and forecasted future occurrences up to 15 days in advance across spatial clusters. This multi-step, cluster-based prediction strategy provides a valuable tool for emergency managers by quantifying where and when demand is likely to increase. In particular, the spatio-temporal visualization of forecasts (Figure 40) enables decision-makers to anticipate critical periods of activity and to allocate personnel, vehicles, and equipment proactively to the most vulnerable areas. Moreover, by linking

predictive outputs to geographical clusters (Figure 41), the model enhances situational awareness by translating numerical forecasts into intuitive spatial risk maps.

The integration of multimodal data sources, including meteorological, temporal, and historical features, proved fundamental to improving interpretability. The TFT model's attention mechanisms made it possible to identify which contextual factors most influenced the predicted outcomes. This level of interpretability extends beyond prediction accuracy: it provides an analytical tool for decision-makers to understand why emergencies are more likely to occur in specific regions and periods, thereby supporting evidence-based planning and operational transparency.

The findings confirm that when properly integrated into decision-support systems, machine learning and statistical models can significantly enhance emergency management practices. By aligning predictive intelligence with real-world operational needs, these models could support emergency dispatchers, planners, and public administrators in making informed, data-driven decisions about resource allocation and risk mitigation.

7 CONCLUSION

Understanding and forecasting emergency events remains a complex challenge that requires robust predictive modeling and the integration of diverse data sources. Throughout this research, we developed and validated a comprehensive framework that addresses this challenge through the combination of spatial segmentation, multimodal data fusion, and advanced time series forecasting using deep learning techniques.

Our results demonstrate that the proposed methodology effectively answers the main research question: well-trained machine learning models, when supported by appropriate data preprocessing and multimodal integration, can generate actionable insights by predicting future emergency events with high reliability. The framework successfully integrates temporal, meteorological, and calendar-based information, and through the use of the TFT, provides not only accurate forecasts but also interpretability, revealing how different variables influence model behavior over time. This transparency enables data-driven decision-making in emergency management, improving resource allocation and operational planning.

Furthermore, by varying prediction horizons and incorporating diverse data modalities, such as weather and holiday information, the framework proved flexible and adaptable across distinct macro-regions. The comparative experiments confirmed that regionalized models outperform generalized ones, highlighting the importance of spatial segmentation for capturing local event dynamics and improving predictive accuracy. These findings reinforce the potential of our approach to support emergency teams in both strategic and operational contexts, from long-term preparedness to short-term response optimization.

FUTURE WORK

Several avenues for future research emerge from this work. First, focusing on single-type event modeling, such as fire incidents, traffic accidents, or medical emergencies, could yield specialized predictors capable of learning fine-grained behavioral patterns. Second, future efforts should explore methods to monitor and mitigate data drift and linkage inconsistencies, ensuring long-term reliability and adaptability of predictive models. Improvements in preprocessing and data quality management will also enhance the robustness of the framework. An additional and particularly relevant opportunity lies in systematically assessing the quality and representativeness of the

data during abnormal periods. For example, the COVID-19 pandemic drastically altered mobility patterns, emergency demands, and operational practices. Dedicated models or separated training phases for such exceptional intervals may yield more accurate predictors and prevent contamination of long-term patterns in the data.

Another promising direction involves integrating the predictive outputs directly into operational dispatcher systems. Embedding predictive analytics into real-time dashboards would allow emergency operators to visualize and act upon forecasts immediately, bridging the gap between model inference and field decision-making. Evaluating such systems in real-world settings would provide valuable insights into their usability, impact on response efficiency, and potential to transform emergency management practices.

In conclusion, the research presented in this thesis demonstrates that machine learning-based predictive systems, when designed under a multimodal and region-aware framework, can meaningfully contribute to anticipating emergency demands and supporting data-driven public safety decisions. The proposed approach establishes a foundation for intelligent, interpretable, and practical forecasting tools that can guide future innovations in smart emergency response systems.

REFERENCES

- AGGARWAL, Charu C. **Data mining: the textbook**. [S.l.]: Springer, 2015.
- AKAIKE, Hirotugu. Fitting autoregressive models for prediction. In: **SELECTED Papers of Hirotugu Akaike**. [S.l.]: Springer, 1969. P. 131–135.
- AKIBA, Takuya; SANO, Shotaro; YANASE, Toshihiko; OHTA, Takeru; KOYAMA, Masanori. Optuna: A next-generation hyperparameter optimization framework. In: **PROCEEDINGS of the 25th ACM SIGKDD international conference on knowledge discovery & data mining**. [S.l.: s.n.], 2019. P. 2623–2631.
- ALKHEDER, Sharaf; TAAMNEH, Madhar; TAAMNEH, Salah. Severity prediction of traffic accident using an artificial neural network. **Journal of Forecasting**, Wiley Online Library, v. 36, n. 1, p. 100–108, 2017.
- ALVES, Luiz GA; MENDES, Renio S; LENZI, Ervin K; RIBEIRO, Haroldo V. Scale-adjusted metrics for predicting the evolution of urban indicators and quantifying the performance of cities. **PloS One**, Public Library of Science San Francisco, CA USA, v. 10, n. 9, e0134862, 2015.
- ALVES, Luiz GA; RIBEIRO, Haroldo V; RODRIGUES, Francisco A. Crime prediction through urban metrics and statistical learning. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 505, p. 435–443, 2018.
- AMEZQUITA-SANCHEZ, JP; VALTIERRA-RODRIGUEZ, M; ADELI, H. Current efforts for prediction and assessment of natural disasters: Earthquakes, tsunamis, volcanic eruptions, hurricanes, tornados, and floods. **Scientia Iranica**, Sharif University of Technology, v. 24, n. 6, p. 2645–2664, 2017.
- BERGMEIR, Christoph; BENÍTEZ, José M. On the use of cross-validation for time series predictor evaluation. **Information Sciences**, v. 191, p. 192–213, 2012. Data Mining for Software Trustworthiness. ISSN 0020-0255. DOI: <https://doi.org/10.1016/j.ins.2011.12.028>. Available from: <https://www.sciencedirect.com/science/article/pii/S0020025511006773>.

- BERGSTRA, James; BARDENET, Rémi; BENGIO, Yoshua; KÉGL, Balázs. Algorithms for hyper-parameter optimization. **Advances in neural information processing systems**, v. 24, 2011.
- BOEING, Geoff. Visual analysis of nonlinear dynamical systems: chaos, fractals, self-similarity and the limits of prediction. **Systems**, MDPI, v. 4, n. 4, p. 37, 2016.
- BOX, George EP; JENKINS, Gwilym M; REINSEL, Gregory C; LJUNG, Greta M. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.
- BROCKWELL, Peter J; DAVIS, Richard A. **Introduction to time series and forecasting**. [S.l.]: Springer, 2002.
- BRUSHLINSKY, NN; AHRENS, Marty; SOKOLOV, SV; WAGNER, Peter. **World fire statistics**. [S.l.]: report, 2017.
- BUITINCK, Lars et al. API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. [S.l.: s.n.], 2013. P. 108–122.
- CAVALCANTE, Rodolfo Carneiro. An adaptive learning system for time series forecasting in the presence of concept drift. Universidade Federal de Pernambuco, 2017.
- CHAMOLA, Vinay; HASSIJA, Vikas; GUPTA, Sakshi; GOYAL, Adit; GUIZANI, Mohsen; SIKDAR, Biplab. Disaster and pandemic management using machine learning: a survey. **IEEE Internet of Things Journal**, IEEE, v. 8, n. 21, p. 16047–16071, 2020.
- CHEN, Ning; LIU, Wenjing; BAI, Ruizhen; CHEN, An. Application of computational intelligence technologies in emergency management: a literature review. **Artificial Intelligence Review**, Springer, v. 52, p. 2131–2168, 2019.
- CHENG, Ching-Hsue; WEI, Liang-Ying. A novel time-series model based on empirical mode decomposition for forecasting TAIEX. **Economic Modelling**, Elsevier, v. 36, p. 136–141, 2014.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, Springer, v. 20, p. 273–297, 1995.

CORTEZ, Bitzel; CARRERA, Berny; KIM, Young-Jin; JUNG, Jae-Yoon. An architecture for emergency event prediction using LSTM recurrent neural networks. **Expert Systems with Applications**, Elsevier, v. 97, p. 315–324, 2018.

COWPERTWAIT, Paul SP; METCALFE, Andrew V. **Introductory time series with R**. [S.I.]: Springer, 2009.

CRYER, Jonathan D. **Time series analysis**. [S.I.]: Duxbury Press Boston, 1986. v. 286.

DEBAS, Haile T; DONKOR, Peter; GAWANDE, Atul; JAMISON, Dean T; KRUK, Margaret E; MOCK, Charles N. **Disease control priorities, (Volume 1): essential surgery**. [S.I.]: World Bank Publications, 2015.

FULLER, Wayne A. Introduction to statistical time series, new york: Johnwiley. **FullerIntroduction to Statistical Time Series1976**, 1976.

GARCIÍA-OLIVARES, Antonio. Substitutability of electricity and renewable materials for fossil fuels in a post-carbon economy. **Energies**, Multidisciplinary Digital Publishing Institute, v. 8, n. 12, p. 13308–13343, 2015.

GHOLAMI, Shahrzad et al. Adversary models account for imperfect crime data: Forecasting and planning against real-world poachers. In: INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS and MULTIAGENT SYSTEMS.

GRAVES, Alex; GRAVES, Alex. **Supervised sequence labelling**. [S.I.]: Springer, 2012.

GREKOUSIS, George; LIU, Ye. Where will the next emergency event occur? Predicting ambulance demand in emergency medical services using artificial intelligence. **Computers, Environment and Urban Systems**, Elsevier, v. 76, p. 110–122, 2019.

GUI, Li; GU, Shen; LU, Feng; ZHOU, Bin; ZHANG, Ling. Prehospital emergency care in Shanghai: present and future. **The Journal of emergency medicine**, Elsevier, v. 43, n. 6, p. 1132–1137, 2012.

HADAD, Dani; AL MASRY, Zeina; NICOD, Jean-Marc; VARNIER, Christophe; ZERHOUNI, Nouredine. A Predictive Decision Approach for Firefighters Using Variable Selection Technique. In: IEEE. 2019 Prognostics and System Health Management Conference (PHM-Paris). [S.l.: s.n.], 2019. P. 258–263.

HÄSKE, David; SCHMÖCKEL, Stefanie. Development of emergency medical services and paramedics. **Notfall & Rettungsmedizin**, v. 25, n. 3, p. 153–155, 2022.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.

HOFFMANN, José Luis Conradi; HORSTMANN, Leonardo Passig; WAGNER, Matheus; VIEIRA, Felipe; DE LUCENA, Mateus Martiénez; FRÖHLICH, Antônio Augusto. Using formal methods to specify data-driven cyber-physical systems. In: IEEE. 2022 IEEE 31st International Symposium on Industrial Electronics (ISIE). [S.l.: s.n.], 2022. P. 643–648.

HUANG, Di; WANG, Shuaian; LIU, Zhiyuan. A systematic review of prediction methods for emergency management. **International Journal of Disaster Risk Reduction**, Elsevier, v. 62, p. 102412, 2021.

HUANG, Jian; CHAI, Junyi; CHO, Stella. Deep learning in finance and banking: A literature review and classification. **Frontiers of Business Research in China**, Springer, v. 14, p. 1–24, 2020.

JALDELL, Henrik; LEBNAK, Prachaksvich; AMORNPETCHSATHAPORN, Anurak. Time is money, but how much? The monetary value of response time for Thai ambulance emergency services. **Value in health**, Elsevier, v. 17, n. 5, p. 555–560, 2014.

KUMAR, D Ashok; MURUGAN, S. Performance analysis of Indian stock market index using neural network time series model. In: IEEE. 2013 international conference on pattern recognition, informatics and mobile engineering. [S.l.: s.n.], 2013. P. 72–78.

LAHAT, Dana; ADALI, Tülay; JUTTEN, Christian. Multimodal data fusion: an overview of methods, challenges, and prospects. **Proceedings of the IEEE**, IEEE, v. 103, n. 9, p. 1449–1477, 2015.

LEE, Ming-Chi. Using support vector machine with a hybrid feature selection method to the stock trend prediction. **Expert Systems with Applications**, Elsevier, v. 36, n. 8, p. 10896–10904, 2009.

LI, Xinde; WANG, Fengyu. A clustering-based evidence reasoning method. **International Journal of Intelligent Systems**, Wiley Online Library, v. 31, n. 7, p. 698–721, 2016.

LI, Yue; LI, Ji; GENG, Jiayu; LIU, Tao; LIU, Xin; FAN, Haojun; CAO, Chunxia. Urban–sub-urban–rural variation in the supply and demand of emergency medical services. **Frontiers in Public Health**, Frontiers Media SA, v. 10, p. 1064385, 2023.

LIM, Bryan; ARIK, Sercan O; LOEFF, Nicolas; PFISTER, Tomas. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. **International Journal of Forecasting**, Elsevier, 2021.

LIN, Lei; WANG, Qian; SADEK, Adel W. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 55, p. 444–459, 2015.

LLOYD, Stuart. Least squares quantization in PCM. **IEEE transactions on information theory**, IEEE, v. 28, n. 2, p. 129–137, 1982.

LU, Tao; DUNYAO, ZHU; LIXIN, Yan; PAN, Zhang. The traffic accident hotspot prediction: Based on the logistic regression method. In: IEEE. 2015 International Conference on Transportation Information and Safety (ICTIS). [S.l.: s.n.], 2015. P. 107–110.

MAJHI, Ritanjali; PANDA, Ganapati; SAHOO, Gadadhar. Efficient prediction of exchange rates with low complexity artificial neural network models. **Expert systems with applications**, Elsevier, v. 36, n. 1, p. 181–189, 2009.

MISHRA, Vipul; AHUJA, Richa; NEZAMUDDIN, N; TIWARI, Geetam; BHALLA, Kavi. Strengthening the capacity of emergency medical services in low and middle income countries using dispatcher-coordinated taxis. **Transportation research record**, SAGE Publications Sage CA: Los Angeles, CA, v. 2674, n. 9, p. 338–345, 2020.

MUKHOPADHYAY, Ayan; PETTET, Geoffrey; SAMAL, Chinmaya; DUBEY, Abhishek; VOROBAYCHIK, Yevgeniy. An online decision-theoretic pipeline for responder dispatch. In: PROCEEDINGS of the 10th ACM/IEEE International Conference on Cyber-Physical Systems. [S.l.: s.n.], 2019. P. 185–196.

MUKHOPADHYAY, Ayan; PETTET, Geoffrey; VAZIRIZADE, Sayyed; VOROBAYCHIK, Yevgeniy; KOCHENDERFER, Mykel; DUBEY, Abhishek. A review of emergency incident prediction, resource allocation and dispatch models. **arXiv preprint arXiv:2006.04200**, 2020.

MUKHOPADHYAY, Ayan; VOROBAYCHIK, Yevgeniy. Prioritized allocation of emergency responders based on a continuous-time incident prediction model. In: INTERNATIONAL Conference on Autonomous Agents and MultiAgent Systems. [S.l.: s.n.], 2017.

MUKHOPADHYAY, Ayan; WANG, Zilin; VOROBAYCHIK, Yevgeniy. A decision theoretic framework for emergency responder dispatch. In: INTERNATIONAL Conference on Autonomous Agents and Multiagent Systems. [S.l.: s.n.], 2018.

MURTAGH, Fionn; STARCK, Jean-Luc; RENAUD, Olivier. On neuro-wavelet modeling. **Decision Support Systems**, Elsevier, v. 37, n. 4, p. 475–484, 2004.

NFPA, NFPA. **Standard for the organization and deployment of fire suppression operations, emergency medical operations, and special operations to the public by career fire departments**. [S.l.]: NFPA, 2020.

NIELSEN, Frank; NIELSEN, Frank. Hierarchical clustering. **Introduction to HPC with MPI for Data Science**, Springer, p. 195–211, 2016.

OLAH, Christopher. **Understanding LSTM Networks**. [S.l.: s.n.], 2015.
<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Accessed: 2024-06-08.

PAL, Sankar K; MITRA, Sushmita. Multilayer perceptron, fuzzy sets, and classification. **IEEE Transactions on neural networks**, Institute of Electrical and Electronic Engineers, v. 3, n. 5, p. 683–697, 1992.

PALIT, Ajoy K; POPOVIC, Dobrivoje. **Computational intelligence in time series forecasting: theory and engineering applications**. [S.l.]: Springer Science & Business Media, 2006.

PAYARES-GARCIA, David; PLATERO, Javier; MATEU, Jorge. A dynamic spatio-temporal stochastic modeling approach of emergency calls in an urban context. **Mathematics**, MDPI, v. 11, n. 4, p. 1052, 2023.

PETTET, Geoffrey; NANNAPANENI, Saideep; STADNICK, Benjamin; DUBEY, Abhishek; BISWAS, Gautam. Incident analysis and prediction using clustering and bayesian network. In: IEEE. 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). [S.l.: s.n.], 2017. P. 1–8.

RDUSSEEUN, LKPJ; KAUFMAN, P. Clustering by means of medoids. In: PROCEEDINGS of the statistical data analysis based on the L1 norm conference, neuchatel, switzerland. [S.l.: s.n.], 1987.

REINHART, Alex. A review of self-exciting spatio-temporal point processes and their applications. **Statistical Science**, JSTOR, v. 33, n. 3, p. 299–318, 2018.

RITCHIE, Hannah; ROSER, Max. Urbanization. **Our World in Data**, 2018.
<https://ourworldindata.org/urbanization>.

ROBERTS, David R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. **Ecography**, v. 40, n. 8, p. 913–929, Aug. 2017. DOI: 10.1111/ecog.02881.

ROQUE, Felipe V.; CAMINHA, Luciano; CECHINEL, Cristian; MARCON, Matheus Z.; KUHNEN, Alex; MUNOZ, Roberto; GRELLERT, Mateus. Learning Analytics for Virtual Industrial Labs: Performance Segmentation and Error Pattern Discovery via Sequential Mining. **IEEE Access**, p. 1–1, 2025. DOI: 10.1109/ACCESS.2025.3631372.

ROQUE, Felipe Vieira; CECHINEL, C.; MUNOZ, R. Analíticas de Aprendizagem Multimodal em Apresentações Orais. In: LUNARDI, Giovani Mendonça; SILVA, Juarez Bento da; SILVA, Isabela Nardi da; SOUZA, Taize Pizoni de (Eds.). **Tecnologia, Inovação e Interdisciplinaridade**. 1. ed. Araranguá: Hard Tech Editora, 2025. v. 1. P. 184–198.

ROQUE, Felipe Vieira; GRELLERT, Mateus; FRÖHLICH, Antônio Augusto. **A LSTM approach to predict emergency events using spatial features**. [S.l.: s.n.], 2022. [Online; accessed <today>]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4261651.

ROUSSEEUW, Peter J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of computational and applied mathematics**, Elsevier, v. 20, p. 53–65, 1987.

SAFAVIAN, S Rasoul; LANDGREBE, David. A survey of decision tree classifier methodology. **IEEE transactions on systems, man, and cybernetics**, IEEE, v. 21, n. 3, p. 660–674, 1991.

SEABOLD, Skipper; PERKTOLD, Josef. statsmodels: Econometric and statistical modeling with python. In: 9TH Python in Science Conference. [S.l.: s.n.], 2010.

SECRETARIAT, UN. Twelfth United Nations Congress on Crime Prevention and Criminal Justice: note/by the Secretariat. < bound method Organization. get_name_with_acronym of< Organization: The ..., 2010.

SEO, Sungyong; CHAN, Hau; BRANTINGHAM, P Jeffrey; LEAP, Jorja; VAYANOS, Phebe; TAMBE, Milind; LIU, Yan. Partially generative neural networks for gang crime classification with partial information. In: PROCEEDINGS of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. [S.l.: s.n.], 2018. P. 257–263.

SHARMA, Dilli Prasad; BEIGI-MOHAMMADI, Nasim; GENG, Hongxiang; DIXON, Dawn; MADRO, Rob; EMMENEGGER, Phil; TOBAR, Carlos; LI, Jeff; LEON-GARCIA, Alberto. Statistical and Machine Learning Models for Predicting Fire and Other Emergency Events in the City of Edmonton. **IEEE Access**, IEEE, 2024.

SHUMWAY, Robert H; STOFFER, David S; STOFFER, David S. **Time series analysis and its applications**. [S.l.]: Springer, 2000. v. 3.

SI, Yain-Whar; YIN, Jiangling. OBST-based segmentation approach to financial time series. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 26, n. 10, p. 2581–2596, 2013.

SMITH, Taylor G. et al. **pmdarima: ARIMA estimators for Python**. [S.l.: s.n.], 2017. [Online; accessed <today>]. Available from: <http://www.alkaline-ml.com/pmdarima>.

STOCK, James; WATSON, Mark; MARCELLINO, Massimiliano. **A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series**. [S.l.], 2005.

STONE, M. Cross-Validatory Choice and Assessment of Statistical Predictions. **Journal of the Royal Statistical Society. Series B (Methodological)**, [Royal Statistical Society, Oxford University Press], v. 36, n. 2, p. 111–147, 1974. ISSN 00359246. Available from: <http://www.jstor.org/stable/2984809>. Visited on: 20 Nov. 2025.

TAIEB, Souhaib Ben. Machine learning strategies for multi-step-ahead time series forecasting. **Universit Libre de Bruxelles, Belgium**, p. 75–86, 2014.

TAIEB, Souhaib Ben; HYNDMAN, Rob J, et al. **Recursive and direct multi-step forecasting: the best of both worlds**. [S.l.]: Department of Econometrics and Business Statistics, Monash Univ., 2012. v. 19.

VIEIRA ROQUE, Felipe; FRÖHLICH, Antônio Augusto; GRELLERT, Mateus. An LSTM approach to predict emergency events using spatial features. **Applied Intelligence**, Springer, v. 55, n. 5, p. 334, 2025.

WANG, Jie; WANG, Jun. Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks. **Neurocomputing**, Elsevier, v. 156, p. 68–78, 2015.

WORSLEY, Marcelo; BLIKSTEIN, Paulo. What's an Expert? Using Learning Analytics to Identify Emergent Markers of Expertise through Automated Speech, Sentiment and Sketch Analysis. In: ERIC. EDM. [S.l.: s.n.], 2011. P. 235–240.

XINGJIAN, SHI; CHEN, Zhourong; WANG, Hao; YEUNG, Dit-Yan; WONG, Wai-Kin; WOO, Wang-chun. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: ADVANCES in neural information processing systems. [S.l.: s.n.], 2015. P. 802–810.

YE, Zhongnan; HSU, Shu-Chien; WEI, Hsi-Hsien. Real-time prediction of structural fire responses: A finite element-based machine-learning approach. **Automation in Construction**, Elsevier, v. 136, p. 104165, 2022.

YUAN, Zhuoning; ZHOU, Xun; YANG, Tianbao. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In: PROCEEDINGS of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. [S.l.: s.n.], 2018. P. 984–992.

ZHANG, Chao; JAIN, Manish; GOYAL, Ripple; SINHA, Arunesh; TAMBE, Milind. Learning, predicting and planning against crime: Demonstration based on real urban crime data. In: AAMAS. [S.l.: s.n.], 2015. P. 1911–1912.

ZHANG, Zhenhua; HE, Qing; GAO, Jing; NI, Ming. A deep learning approach for detecting traffic accidents from social media data. **Transportation research part C: emerging technologies**, Elsevier, v. 86, p. 580–596, 2018.

ZHAO, Xiangyu; TANG, Jiliang. Exploring transfer learning for crime prediction. In: IEEE. 2017 IEEE International Conference on Data Mining Workshops (ICDMW). [S.l.: s.n.], 2017. P. 1158–1159.

APPENDIX A – PUBLICATIONS SUMMARY

This chapter summarizes the scientific publications produced during the Ph.D. research. The contributions reflect the evolution of the work and correspond to various stages of the framework presented in this thesis, including foundational modeling, spatial-temporal forecasting, multimodal data integration, and applications in learning analytics.

A.1 JOURNAL ARTICLES

1. *An LSTM approach to predict emergency events using spatial features*. Applied Intelligence. (VIEIRA ROQUE et al., 2025)
This article addresses spatial segmentation and LSTM modeling strategies.
2. *A LSTM approach to predict emergency events using spatial features*. (preprint SSRN) (ROQUE, Felipe Vieira et al., 2022)
Early work that laid the foundation for temporal forecasting approaches.
3. *Learning Analytics for Virtual Industrial Labs: Performance Segmentation and Error Pattern Discovery via Sequential Mining*. (ROQUE, Felipe V. et al., 2025).
Applies sequence mining to analyze learner performance patterns in remote laboratories.

A.2 CONFERENCE PAPERS AND OTHER PUBLICATIONS

1. *Using formal methods to specify data-driven cyber-physical systems*. (HOFFMANN et al., 2022)
Collaborative work extending formal specification ideas into data-driven systems.
2. *Analíticas de Aprendizagem Multimodal em Apresentações Orais*. In: Tecnologia, Inovação e Interdisciplinaridade. 1st Ed. Araranguá: UFSC. (ROQUE, Felipe Vieira et al., 2025)
Book chapter exploring multimodal learning analytics in educational contexts.

APPENDIX B – COMPLEMENTARY RESULTS

B.1 CASE STUDY 1

B.1.1 Additional vizualizations

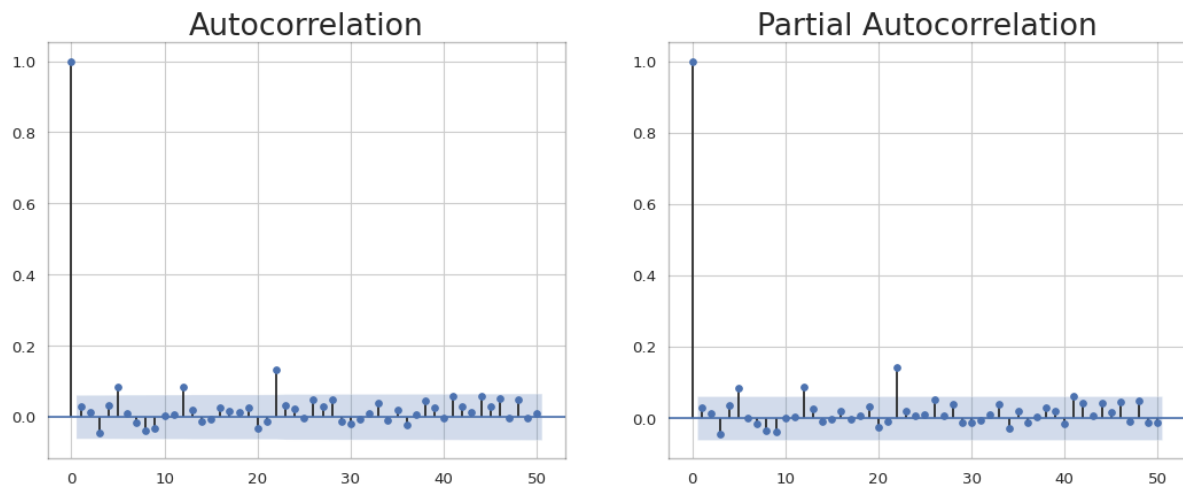


Figure 42 – ACF and PACF plots for DS-Truck occurrences.

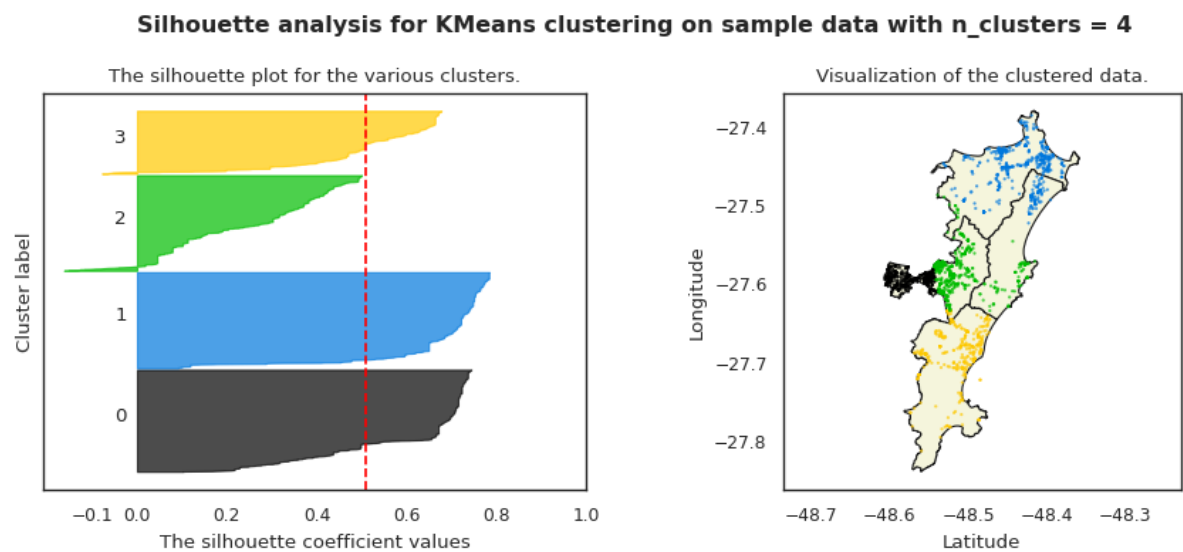


Figure 43 – Best k value for the problem and its visual separation seem from silhouettes for DS-Truck.

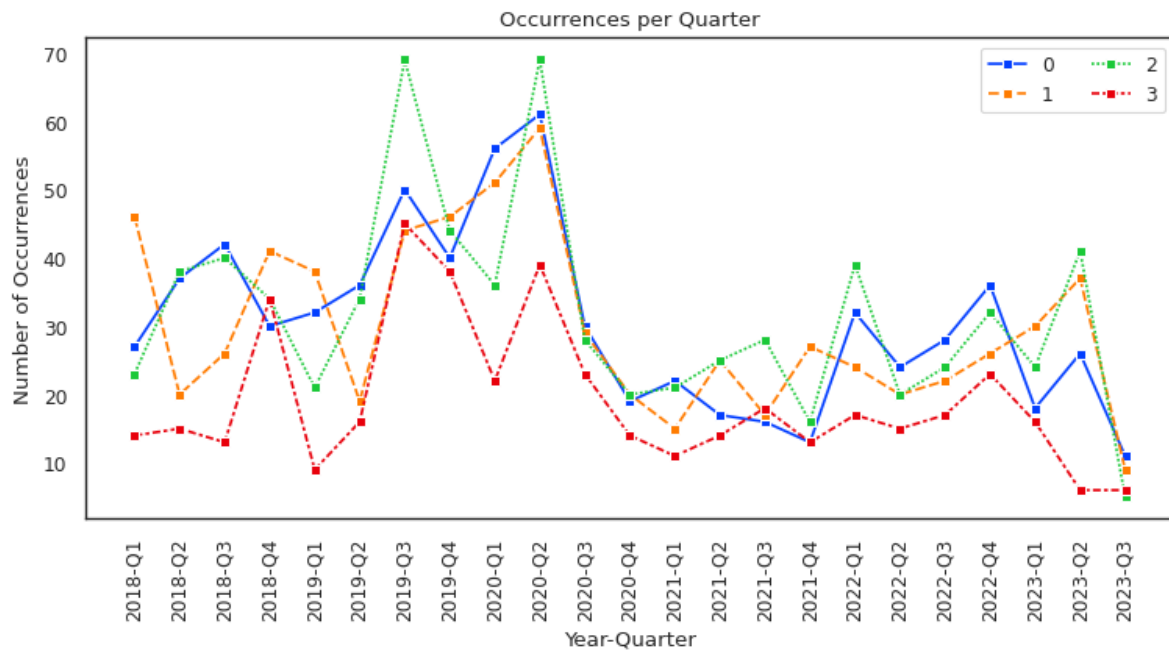


Figure 44 – Four clusters for DS-Truck events over time divided by the clustering technique with move average for 30-day.

B.1.2 Additional tables

Table 22 – Comparing retrained LSTM regression results for different look ahead values (direct forecasting).

Days Ahead	DS-Ambulance			DS-Truck		
	MAE	MSE	RMSE	MAE	MSE	RMSE
1	1.29	3.02	1.73	0.71	1.03	1.01
2	1.68	2.15	4.65	0.79	1.10	1.22
3	1.60	2.06	4.24	0.75	1.09	1.19
4	1.69	2.19	4.82	0.78	1.11	1.24
5	1.61	2.08	4.36	0.77	1.10	1.21
6	1.66	2.14	4.59	0.77	1.08	1.18
7	1.68	2.17	4.72	0.78	1.10	1.22
8	1.64	2.13	4.55	0.77	1.07	1.16
9	1.61	2.07	4.31	0.76	1.07	1.16
10	1.63	2.10	4.42	0.75	1.07	1.16

Table 23 – Comparing LSTM regression results for a different look ahead by feeding back the input (recursive forecasting).

Days Ahead	DS-Ambulance			DS-Truck		
	MAE	MSE	RMSE	MAE	MSE	RMSE
1	1.29	3.02	1.73	0.71	1.03	1.01
2	1.59	2.03	4.15	0.73	1.07	1.15
3	1.59	2.04	4.17	0.72	1.07	1.15
4	1.58	2.03	4.14	0.72	1.07	1.14
5	1.58	2.03	4.12	0.72	1.07	1.14
6	1.57	2.02	4.10	0.71	1.07	1.14
7	1.57	2.02	4.08	0.71	1.07	1.14
8	1.57	2.02	4.09	0.71	1.07	1.14
9	1.57	2.02	4.09	0.71	1.07	1.15
10	1.57	2.02	4.11	0.71	1.07	1.15

B.2 CASE STUDY 2

Silhouette analysis for KMeans clustering on sample data with $n_clusters = 7$

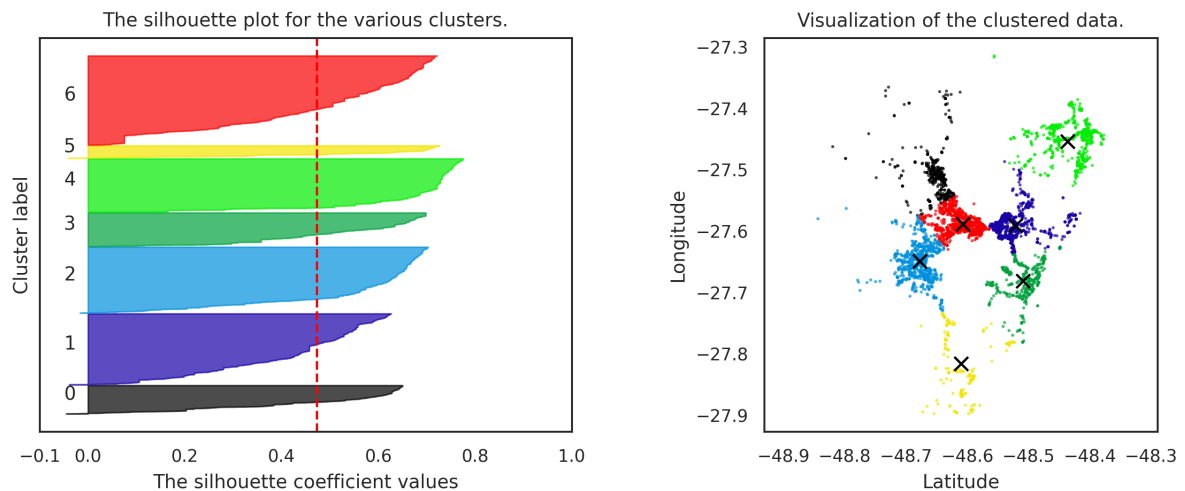


Figure 45 – Best k value for the problem and its visual separation seem from silhouettes for DS-Truck.

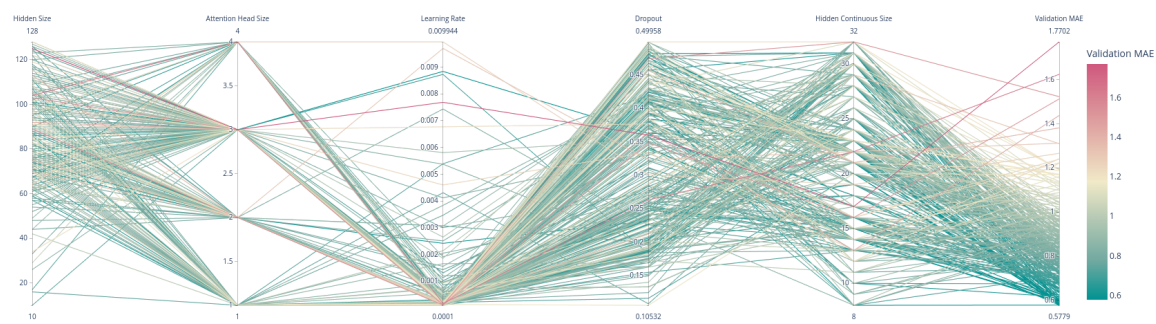


Figure 46 – Four clusters for DS-Truck events over time divided by the clustering technique with move average for 30-day.