Article

Preserving Geo-Indistinguishability of the Emergency Scene to Predict Ambulance Response Time

Héber H. Arcolezi [1]*, Selene Cerna [1], Christophe Guyeux [1]*, and Jean-François Couchot [1]

[1] Femto-ST Institute, Univ. Bourgogne Franche-Comté, UBFC, CNRS, Belfort, France; {heber.hwang_arcolezi, selene_leya.cerna_nahuis, christophe.guyeux, jean-francois.couchot}@univ-fcomte.fr

* Correspondence: heber.hwang_arcolezi@univ-fcomte.fr (H.H.A), christophe.guyeux@univ-fcomte.fr (C.G)

Abstract: Emergency medical services (EMS) provide crucial emergency assistance and ambulatory services. One key measurement of EMS's quality of service is their ambulances' response time (ART), which generally refers to the period between EMS notification and the moment an ambulance arrives on the scene. Due to many victims require care within adequate time (*e.g.*, cardiac arrest), improving ARTs is vital. This paper proposes to predict ARTs using machine learning (ML) techniques, which could be used as a decision-support system by EMS to allow a dynamic selection of ambulance dispatch centers. However, one well-known predictor of ART is the location of the emergency (*e.g.*, if it is urban or rural areas), which is *sensitive data* because it can reveal who received care and for which reason. Thus, we considered the 'input perturbation' setting in the privacy-preserving ML literature, which allows EMS to sanitize each location data independently and, hence, ML models are trained only with sanitized data. In this paper, geo-indistinguishability was applied to sanitize each emergency location data, which is a state-of-the-art formal notion based on differential privacy. To validate our proposals, we used retrospective data of an EMS in France, namely, Departmental Fire and Rescue Service of Doubs, and publicly available data (*e.g.*, weather and traffic data). As shown in the results, the sanitization of location data and the perturbation of its associated features (*e.g.*, city, distance) had no considerable impact on predicting ARTs. With these findings, EMSs may prefer using and/or sharing sanitized datasets to avoid possible data leakages, membership inference attacks, or data reconstructions, for example.

Keywords: Emergency medical services; Emergency medicine; Decision support system; Prehospital emergency care; Ambulance response time; Machine learning; Geo-indistinguishability; Differential privacy; Privacy-preserving machine learning; Input perturbation.

Citation: Arcolezi, H.H.; Cerna, S.; Guyeux, C; Couchot, J.F. Preserving Geo-Indistinguishability of the Emergency Scene to Predict Ambulance Response Time. *Math. Comput. Appl.* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Ambulance response time (ART) is a key component for evaluating pre-hospital emergency medical services (EMS) operations. ART refers to the period between the notification and the moment an ambulance arrives at the emergency scene [1,2], and it is normally divided into two periods: the pre-travel delay, from the notification to the ambulance dispatch, and the travel time, from the ambulance dispatch to arrival on-scene. In many urgent situations (*e.g.*, cardiovascular emergencies, trauma, or respiratory distress), the victims need first-aid treatment within adequate time to increase survival rate [1–6] and, hence, improving ART is vital.

In many parts of the world, such as France, fire departments are responsible for many critical situations, including fires, hazards, severe storms, floodings, as well as non-urgent and urgent EMS calls (*e.g.*, traffic accidents, drowning). In this paper, we analyzed EMS operations of the Departmental Fire and Rescue Service of Doubs (SDIS 25), which has 71 centers currently deployed across the Doubs region in France to attend to its population. As noticed in [7,8], the SDIS 25 and fire departments in general, have been facing a continuous increase in the number of interventions over the years, which may have adverse consequences on ARTs. For instance, the pre-travel delay affects directly ARTs if there is a lack of human and material resources when a call is received. This means, if there is a lack of firefighters, ambulances, or both, ART may be higher than allowed and, hence, a breakdown in the SDIS 25 service occurs [9]. This inability to assist within the time limits impacts negatively both EMS and victims because the safety of a certain area or population will be at risk. Thus, there is a need for an intelligent ART prediction system, which can assist SDIS 25 (and EMS, in general) in the dispatching of ambulances.

Indeed, predicting ART is useful for many reasons. First of all, it can help in choosing the best center to provide the ambulance. At present, for SDIS 25, each city in the department is associated with an ordered list of centers with the needed engine to respond, so that the first centers are the most likely to provide a rapid and adequate response. This sorting is guided both by the travel time needed to get from the center to the city and by the material and human resources of the center. However, this ordering of centers by city is fixed once and for all. While it does take into account the actual distance of the travel (not the Euclidean distance, for example), it does not consider the state of the road traffic, weather conditions, etc. This way, a center might be a little closer to the emergency scene than another, but may occasionally have a longer travel time due to traffic congestion. Predicting ART would therefore make it possible to move from static center scheduling to dynamic scheduling. It would also make it possible to estimate the exit time of vehicles partially and to see in advance whether, at a given moment, a center is at risk of running out of ambulances. In other words, it enables the anticipation of breakdowns and the redeployment of resources. Lastly, in the long term, it can be an element of a simulator to determine the evolution of response time and breakdowns during the creation or relocation of a center, the modification of resources by the center, etc.

As aforementioned, an important factor of ART is the *location* of the intervention, *e.g.*, in dense urban areas, the distance may be short, but the travel time may be longer due to traffic congestion. On the other hand, travel distance and travel time may be longer for rural areas. In other words, the location information is of great importance for the prediction of travel time and, naturally, ART [10,11]. However, the location of an emergency is also regarded as *sensitive data* because it can reveal who received care and for which reason. For example, by knowing that one intervention took place in front of the house of a debilitated person, attackers with auxiliary information may accurately infer that this person received care and (mis)use this information for their own good. Indeed, location privacy is an emerging and active research topic in the literature [12–14] as publicly exposing users' location raises major privacy issues. A common way to achieve location privacy is by applying a *location obfuscation* mechanism. In [14], the authors proposed geo-indistinguishability (GI), which is based on the state-of-the-art differential privacy (DP) [15] model, to protect the location privacy of users. GI has received considerable attention due to its effectiveness and simplicity of implementation (*e.g.*, Location Guard [16]).

In this paper, we propose to sanitize, independently, each emergency location data with GI before training any ML techniques to predict ARTs. In our context, besides the own location, with the exact coordinates of both SDIS 25 centers and the emergency scenes, one can retrieve important features such as the distance and estimated travel time. However, if the location is sanitized via GI, many other explanatory variables (*e.g.*, distance, travel time, city) would be 'perturbed' too. In the privacy-preserving data mining literature, training ML models with sanitized data is common practice [7,17–22], which is also known as *input perturbation* [23]. Different from objective [24] and gradient [25] perturbation settings, input perturbation is the easiest method to apply and it is independent of any ML and post-processing techniques. We also remark that input perturbation is in accordance with real-world applications where EMS would only use and/or share sanitized data with trusted third parties to *train* and develop ML-based decision support systems.

To summarize, this paper proposes the following contributions:

- Recognize the most influential variables when building accurately ML-based models to predict ART. This would allow other EMS to collect these variables and recreate our methodology or develop their own considering their policies.
- Evaluate the effectiveness of several values of ϵ (*i.e.*, the privacy budget), to sanitize emergency location data with GI and train ML-based models to predict ART. To the author's knowledge, this is the first work to assess the impact of geo-indistinguishability on sanitizing the location of emergency scenes when training ML models for such an important task. While predicting ART is a means to allow EMS to save more lives, we notice that it is also possible to do so while preserving the victims' privacy.

Outline: The remainder of this paper is organized as follows. In Section 2, we describe the material and methods used in this work, *i.e.*, the geo-indistinguishability privacy notion that we are considering, the data presentation (context, collection, and analysis), the sanitization of emergency scenes with GI, the ML models, and the experimental setup. In Section 3, we present the results of our experiments and our discussion. Lastly, in Section 4, we present the concluding remarks and future directions.

2. Materials and Methods

In this section, we revise the notion of privacy considered in this paper, namely, geo-indistinguishability (Subsection 2.1), we provide a description of the processing of interventions by SDIS 25 (Subsection 2.2), the data collection process (Subsection 2.3), the analysis of SDIS 25 ARTs (Subsection 2.4), the GI-based sanitization of emergency location data (Subsection 2.5), the ML models used for predicting ARTs (Subsection 2.6), and the experimental setup (Subsection 2.7).

2.1. Geo-indistinguishability

Differential privacy [15] has been accepted as the *de facto* standard for data privacy. DP was developed in the area of statistical databases but it is now applied to several fields. Furthermore, DP has also been extended to a local model (*a.k.a.* LDP [23]) in which users sanitize their data before sending it to the server. While DP is well-suited to the case of trusted curators, with LDP, users do not need to trust the curator.

Geo-indistinguishability [14] is based on a generalization of DP developed in [26] and has been proposed for preserving location privacy without the need of a trusted curator (*e.g.*, a malicious location-based service – LBSs). A mechanism satisfies ϵ-GI if for any two locations x_1 and x_2 within a radius r, the output y of them is (ϵ, r)-geo-indistinguishable if we have:

$$\frac{\Pr(y|x_1)}{\Pr(y|x_2)} \leq e^{\epsilon r}, \forall r > 0, \forall y, \forall x_1, x_2 : d(x_1, x_2) \leq r.$$

Intuitively, this means that for any point x_2 within a radius r from x_1, GI forces the corresponding distributions to be at most $l = \epsilon r$ distant. In other words, the level of distinguishability l increases with r, *e.g.*, an attacker can distinguish that the user is in Paris rather than London but can hardly (controlled by ϵ) determine the user's exact location. Although both GI and DP use the notation of ϵ to refer to the privacy budget, they cannot be compared directly because ϵ in GI contains the unit of measurement (*e.g.*, meters).

On the continuous plane (as we consider in this paper), an intuitive polar Laplace mechanism has been proposed in [14] to achieve GI, which is briefly described in the following. Rather than reporting the user's true location $x \in \mathbb{R}^2$, we report a point $y \in \mathbb{R}^2$ generated randomly according to $D_\epsilon(y) = \frac{\epsilon^2}{2\pi} e^{-\epsilon d_2(x,y)}$. Algorithm 1 shows the pseudocode of the polar Laplace mechanism in the continuous plane. More specifically, the noise is drawn by first transforming the true location x to polar coordinates. Then, the angle θ is drawn randomly between $[0, 2\pi)$ (line 3), and the distance r is drawn from $C_\epsilon^{-1}(p)$ (line 5), which is calculated using the negative branch W_{-1} of the Lambert W function. Finally, the generated distance and angle are added to the original location.

Algorithm 1 Polar Laplace mechanism in continuous plane [14]

1: **Input :** $\epsilon > 0$, real location $x \in \mathbb{R}^2$.
2: **Output :** sanitized location $y \in \mathbb{R}^2$.
3: Draw θ uniformly in $[0, 2\pi)$
4: Draw p uniformly in $[0, 1)$
5: Set $r = C_\epsilon^{-1}(p) = -\frac{1}{\epsilon}\left(W_{-1}\left(\frac{p-1}{e}\right) + 1\right)$
6: **return :** $y = x + \langle r\cos(\theta), r\sin(\theta)\rangle$

2.2. Process Flow description

The Departmental Fire and Rescue Service of Doubs currently has 71 centers deployed throughout the region of Doubs, France, serving a population of around 540,000 people. The focus of this paper is on interventions with *victims* that were further transported to hospitals. In these interventions, there was a need for an *emergency and victim assistance vehicle* (*a.k.a.* Véhicule de Secours et d'Assistance aux Victimes - VSAV). VSAVs are equipped with adequate material and personnel for first-aid treatment in urgent situations. In this paper, we interchangeably use the term 'ambulance' when referring to VSAV.

The process of an intervention is briefly described in the following. First, an emergency call is received and treated by an operator. Next, the adequate crew/engine is notified (t_1). Once the sufficient armament is gathered, the ambulance goes to the emergency scene (t_2). Upon arriving on-scene, the crew uses a mechanical system to report their arrival (t_3). We focus on the ART period, which is calculated as: $ART = t_3 - t_1$.

The operation process to decide the adequate SDIS 25 center to attend the intervention depends on the exact *location* of the intervention. As stated previously, there is a city, a district, and a zone that jointly define a list of priority centers, which are responsible for the call. The reason for such a list is because a single center may not have sufficient resources at time t_1 to attend an intervention. In this case, if the first center of the list does not have sufficient resources, another center(s) would be in charge of the call. Also, many situations may generate several victims (*e.g.*, traffic accidents, floods). In these cases, a single intervention can require more than one ambulance, which can come from different centers depending on the availability of resources. This means different ARTs for the same intervention and, therefore, we focus on each ambulance in our analysis and predictions.

In addition, although in some countries the *reason* of the emergency may require a recommended ART [27,28], for SDIS 25, ART depends on the *Zone* as detailed in [9]. There are three zones: Z1 refers to urban areas, Z2 refers to semi-urban areas, and Z3 refers to rural ones. Therefore, SDIS 25 ambulances should arrive on-scene with $ART \leq 10$ minutes (min) on Z1 and with $ART \leq 25$ min on Z2 and Z3, *i.e.*, including the pre-travel delay (gathering armament) and travel-time. If these time limits are not reached, a breakdown in SDIS 25 services is generated [9] and the victim state may be negatively impacted [1,5]. Lastly, SDIS 25 may also help other EMS outside the Doubs region, and in this case, there is no pre-defined ART limit by SDIS 25.

2.3. Data collection

We used retrospective data of EMS operations recorded by SDIS 25. All interventions with *victim*, that were attended by SDIS 25 centers with a VSAV, were eligible for inclusion. These data covered the period of January 2006 to June 2020. The main attributes of these data are described in the following:

- *ID* is a unique identifier for each intervention;
- *SDate* is the time SDIS 25 took charge of the intervention by processing the call;
- *ADate* is the time when an ambulance arrived on the emergency scene;
- *Center* is the SDIS 25 center from which the ambulance left;
- *Location* is the precise location (latitude, longitude) of the intervention;
- *Zone* is either urban (Z1), semi-urban (Z2), or rural (Z3);
- *City* is the municipality where the intervention took place. A city may have zero or more *Districts*.

Each ambulance represents one sample, *i.e.*, a single intervention may have received one or more ambulances. The ART variable was calculated as $ART = ADate - SDate$. We excluded outlying observations with ART of less than 1 minute and with ART of more than 45 minutes, which represented less than 1.4% of the original number of samples.

Using *SDate*, we have added temporal information such as: year, month, day, weekday, hour, and categorical indicators to denote holidays, end/start of the month, and end/start of the year. Besides, with the exact coordinates from both *Center* and emergency's *Location*, we calculated the great-circle distance[1] to add as a feature, which is the shortest distance between two points on the surface of a sphere. We have added the number of interventions in the past hour and the number of active interventions in the current hour. As also remarked in the literature [3,10], the number of interventions on previous hours might impact ART. In addition, external data that may affect ART were gathered from the following sources:

- Bison-Futé [29] provides prediction of traffic level for the Doubs region as indicators ranging from 1 (regular flow) to 4 (extremely difficult flow) per day. We added these indicators according to *SDate*;
- Météo-France [30] supplies historical weather information such as precipitation, temperature, wind speed, and gust speed. We added weather data per hour according to *SDate*;
- OSRM API [31] gives the driving distance on the fastest route and its travel time duration. This way, with the coordinates from both *Center* and emergency's *Location*, we added these two features, *i.e.*, estimated travel time in minutes and driving distance in kilometers (km), for each ambulance.

2.4. Data analysis

After removing outlying observations, the dataset at our disposal has 186, 130 dispatched ambulances from SDIS 25 centers that attended 182, 700 EMS interventions. The frequency on the number of dispatched ambulances per zone is 39.62% (Z1), 33.38% (Z2), 26.71% (Z3), and 0.29% (outside the Doubs region), respectively. Figure 1 illustrates the distribution of our variable of interest, namely ART, via three histograms with bins of 1 minute for each zone within the Doubs region. One can notice that the ART distributions follow a typical right-skewed distribution also observed in other works/countries [3,27,32]. The mean and standard deviation (std) values for zones Z1, Z2, and Z3 are 8.79 ± 5.66 min, 11.43 ± 6.15 min, and 15.38 ± 6.41 min, respectively. SDIS 25 had about 79.52% of the time $ART \leq 10$ min on zone Z1, and had about 95.76% and 92.50% of the time $ART \leq 25$ min on zones Z2 and Z3, respectively.

Figure 2 illustrates the hourly number of dispatched ambulances (left-hand plot) and the cumulative ART in hours per day of the week and hour in the day (right-hand plot). One can notice that the number of dispatched ambulances is notably related to the hour in the day, *i.e.*, there were more interventions in working periods rather than between 0h to 6h. This behavior is also noticed in the works [11,27]. Besides, from 8h in

[1] https://en.wikipedia.org/wiki/Great-circle_distance

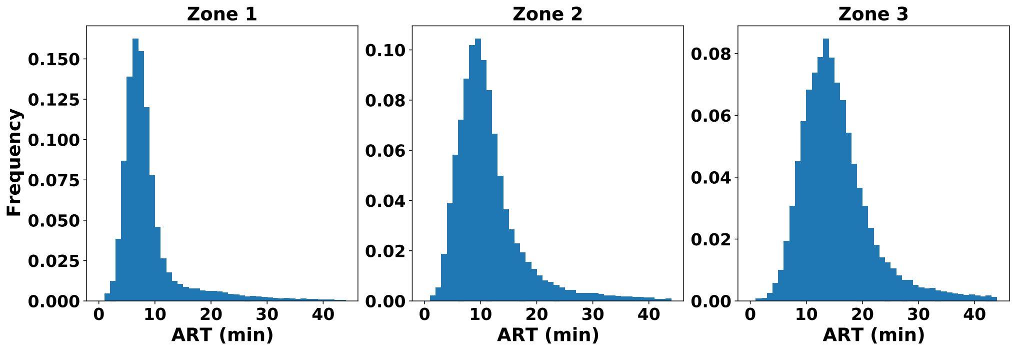

Figure 1. Distribution of the ART variable for zones Z1, Z2, and Z3, respectively.

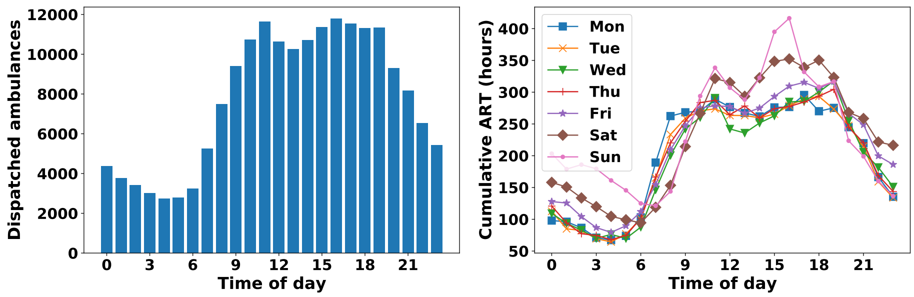

Figure 2. Histogram of the number of dispatched ambulances per hour in the day (left-hand plot) and cumulative ART in hours per day of the week and hour in the day (right-hand plot).

the morning on, the ART starts to increase and remains high up to 19h when it starts to decrease. For instance, ambulances dispatched during working periods are more likely to traffic congestion and, naturally, to undergo through longer travel time. Secondly, due to the number of interventions in a given hour, SDIS 25 centers may have taken more time to dispatch ambulances if their resources were in use in other incidents. A slightly different profile can be seen on weekends, with noticeable higher cumulative ARTs in the late night (0-6h) and during some hours of the day too.

Summary statistics per year and per zone are shown in Table 1. The metrics in this table includes the total number of dispatched ambulances (Nb. Amb.), and descriptive statistics such as mean and standard deviation (std) values for the ART variable. We recall that for the year 2020, these statistics are up to June 2020 only. As also noticed in [7,8], the number of interventions increases throughout the years. The year 2010 presented high values in comparison with all other years, *e.g.*, for Z1, the average ART was above the 10 min recommendation.

2.5. Preserving emergency location privacy with geo-indistinguishability

To preserve the privacy of each emergency scene, we apply the polar Laplace mechanism in Alg. 1 to the *Location* attribute of each intervention. This means, even if our dataset is per ambulance dispatch (*i.e.*, 186,130 ambulances), we used the same sanitized value per intervention (*i.e.*, 182,700 unique interventions). Although in [14] the authors propose two further steps to Alg. 1, *i.e.*, discretization and truncation, both steps can be neglected in our context. This is, first, because SDIS 25 may also help other EMS outside the Doubs region as we discussed in Subsection 2.2, and second, we assume that any location in the continuous plane can be an emergency scene. While reporting an approximate location in the middle of a river may not have much sense in LBSs, in an emergency dataset with approximate locations, this may indicate an urgency for someone who drowned in the river, for example.

We used five different levels for the privacy budget $\epsilon = l/r$, where l is the privacy level we want within a radius r. Table 2 exhibits the five different levels of privacy. For

Year	**Z1**			**Z2**			**Z3**		
	Nb. Amb.	Mean	Std	Nb. Amb.	Mean	Std	Nb. Amb.	Mean	Std
2006	197	9.23	4.41	367	11.25	5.50	354	14.27	5.40
2007	236	7.39	3.05	671	10.79	5.04	595	14.35	5.52
2008	799	8.69	6.04	1,055	11.19	5.32	911	14.53	6.02
2009	1,363	8.76	6.05	2,087	11.08	5.67	1,872	14.94	6.46
2010	2,643	10.08	7.23	2,797	12.48	6.85	2,483	16.01	7.22
2011	5,971	8.26	5.61	4,276	11.24	6.13	3,295	14.50	6.25
2012	6,078	8.66	5.89	4,661	11.18	6.39	3,602	14.86	6.24
2013	6,780	8.82	5.72	5,048	11.03	6.11	3,972	15.07	6.30
2014	6,847	8.37	5.23	5,481	10.80	5.86	4,240	14.91	6.34
2015	7,226	8.46	5.50	5,596	10.86	5.78	4,643	15.02	6.12
2016	7,510	8.50	5.35	6,179	11.19	5.92	4,861	15.32	6.35
2017	8,650	8.76	5.32	7,251	11.49	6.01	5,523	15.51	6.36
2018	9,051	8.90	5.46	7,641	11.64	6.11	5,956	15.59	6.23
2019	7,030	9.42	6.02	6,238	12.29	6.66	5,016	16.60	6.88
2020*	3,397	9.73	5.87	2,843	12.59	6.56	2,449	16.46	6.44

Table 1: Mean and std values for the ART variable and the total number of dispatched ambulances (Nb. Amb.) per year in zones Z1, Z2, and Z3, respectively. For 2020, we only consider cases of the first semester.

the sake of illustration, Figure 3 exhibits three maps of the Doubs region with the points of original raw location (left-hand plot), $\epsilon = 0.005493$-GI location (middle plot), and $\epsilon = 0.002747$-GI location (right-hand plot). As one can notice, with an intermediate privacy level ($l = \ln(3), r = 400$), locations are more spread throughout the map while with a lower privacy level ($l = \ln(3), r = 200$), locations approximate the real clusters.

$\epsilon = l/r$	l	r (meters)
0.005493	$\ln(3)$	200
0.002747	$\ln(3)$	400
0.001155	$\ln(2)$	600
0.000866	$\ln(2)$	800
0.000693	$\ln(2)$	1,000

Table 2: Values of $\epsilon = l/r$ for sanitizing emergency location data with GI.

With the new *Location* values of each intervention, we also reassigned the city, the district, and the zone when applicable. In addition, we recalculated the following features associated with it: the great-circle distance, the estimated driving distance, and estimated travel time. The two features recalculated with OSRM API only consider roads, *i.e.*, if the obfuscated location is in the middle of a farm, the closest route estimates the driving distance and travel time until the closest road. We also highlight that if the new coordinates of the emergency scene indicate a location closer to another SDIS 25 center, even in real life, it would not imply that this center took charge of the intervention. Therefore, the *center* attribute was not 'perturbed'.

To show the impact of the noise added to the *Location* attribute, Table 3 exhibits the percentage of time that categorical attributes (zone, city, and district) were 'perturbed' (*i.e.*, reassigned); the mean and std values of the great-circle distance attribute and its correlation with the ART variable (Corr. ART). In Table 3, we report the mean(std) values since we repeated our experiments with 10 different seeds (*i.e.*, DP algorithms are randomized). Although we did not include the estimated driving distance and estimated travel time from OSRM API in this analysis, in preliminary tests, we noticed that these two features follow a similar pattern as the great-circle distance attribute.

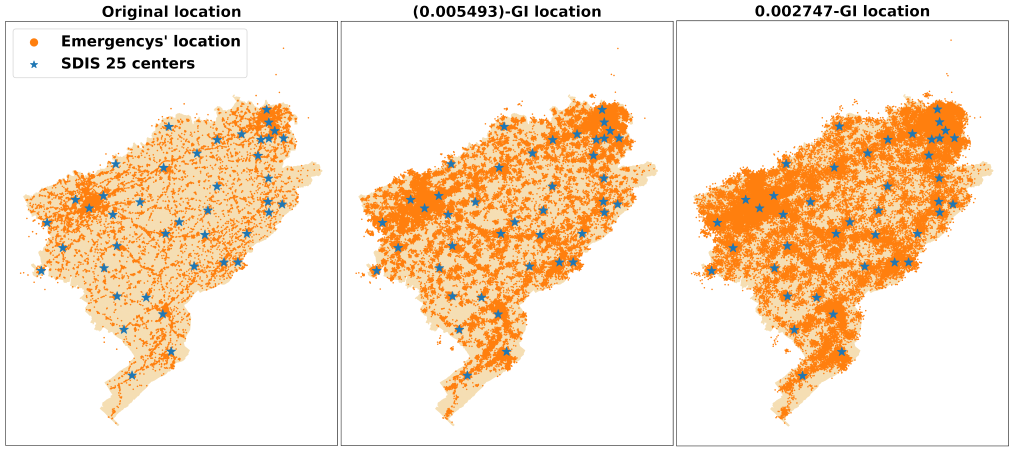

Figure 3. Emergency locations and SDIS 25 centers throughout the Doubs region: original data (left-hand plot), $\epsilon = 0.005493$-GI data (middle plot), and $\epsilon = 0.002747$-GI data (right-hand plot).

From Table 3, one can notice that many features are perturbed due to sanitization of emergency's location with GI. With high levels of ϵ (*i.e.*, less private), the city and the zone suffer low 'perturbation'. On the other hand, district is reassigned many times as it is geographically smaller than the others. When $\epsilon = 0.000866$, the city is already reassigned more than 50% of the time and the district about 75% of the time. Moreover, one can notice that the mean and std values of the great-circle distance increase as the ϵ parameter decreases (*i.e.*, more private). Because $\epsilon = l/r$, making l smaller and/or r higher, the stricter ϵ becomes, and therefore more noise is added to the original locations. Besides, the correlation between the great-circle distance with the ART variable decreases proportionally as ϵ becomes smaller.

Data	**Zone**	**City**	**District**	**Great-circle Dist. (km)**		
	'Perturbation' (%)			Mean	std	Corr. ART
Original	-	-	-	3.44	3.72	0.369
$\epsilon = 0.005493$	5.20(0.05)	7.68(0.06)	25.8(0.05)	3.48(1e-3)	3.72(7e-4)	0.367(2e-4)
$\epsilon = 0.002747$	11.3(0.05)	17.6(0.10)	41.5(0.12)	3.57(1e-3)	3.72(1e-3)	0.362(2e-4)
$\epsilon = 0.001155$	28.1(0.06)	42.3(0.10)	66.2(0.09)	4.03(3e-3)	3.74(3e-3)	0.335(5e-4)
$\epsilon = 0.000866$	35.5(0.10)	52.4(0.11)	74.0(0.11)	4.38(3e-3)	3.81(4e-3)	0.313(1e-3)
$\epsilon = 0.000693$	41.4(0.12)	60.3(0.09)	79.4(0.05)	4.77(6e-3)	3.92(5e-3)	0.288(1e-3)

Table 3: Percentage of perturbation for categorical attributes (city, zone, and district) according to ϵ and statistical properties (mean and std values and correlation with ART) of the original and GI-based datasets for the great-circle distance attribute. Mean(std) values are reported since we repeated our experiments with 10 different seeds.

2.6. Machine learning models

Four state-of-the-art ML techniques have been considered during our experiments, to predict the scalar ART outcome in a regression framework. They are briefly described in the following:

- Extreme Gradient Boosting (XGBoost) [33] is a decision-tree-based ensemble ML algorithm that produces a forecast model based on an ensemble of weak forecast models (decision trees). XGBoost uses a novel regularization approach over standard gradient boosting machines, which significantly decreases model's complexity. The system is optimized by a quick parallel tree construction and adapted to be fault-tolerant under distributed environments.
- Light Gradient Boosted Machine (LGBM) [34] is a novel gradient boosting framework, which implemented a leaf-wise strategy. This strategy significantly reduces

computational speed and resource consumption in comparison to other decision tree-based algorithms.

- Multilayer Perceptron (MLP) is an artificial neural network of the feedforward type [35]. These algorithms are based on the interconnection of several units (neurons) to transmit signals, which are normally structured into three or more layers, input, hidden(s), and output. We used the Keras library [36] to implement our deep learning models.
- Least Absolute Shrinkage and Selection Operator (LASSO), a method of contracting the coefficients of the regression, whose ability to select a subset of variables is due to the nature of the constraint on the coefficients. Originally proposed by Tibshirani [37] for models using the standard least squares estimator, it has been extended to many statistical models such as generalized linear models, etc. We used the LASSO implementation from the Scikit-learn library [38].

2.7. Experiments

Because in Table 3 there are low variations (*i.e.*, small std values) considering 10 different executions on all analyzed features, we ran our experimental validation only once. In our experiments, each sample corresponds to one ambulance dispatch, in which we included temporal features (*e.g.*, hour, day), weather data (*e.g.*, pressure, temperature), traffic data, the emergency's location (latitude and longitude in radians), and computable features (*e.g.*, distance, travel time). The scalar target variable is the ART in minutes, which is the time measured from the EMS notification to the ambulance's arrival on-scene. All numerical features (*e.g.*, temperature) were standardized using the *StandardScaler* function from the Scikit-Learn library. Categorical features (*e.g.*, center, zone, hour) were encoded using mean encoding, *i.e.*, the mean value of the ART variable with respect to each feature. The target variable, namely ART, was kept in its original format (minutes) since no remarkable improvement was achieved with scaling.

We divided our dataset into training (years 2006-2019) and testing (six months of 2020) sets to evaluate our models. Thus, five models per ML technique (*i.e.*, XGBoost, LGBM, MLP, and LASSO) were built to predict ART on each month of 2020 considering the sanitized datasets with different levels of ϵ-GI location data (*cf.* Table 2). In addition, for comparison, we also trained one additional model per ML technique with original raw data. All models were trained continuously, *i.e.*, at the end of each month, the new known data were added to the training set. Lastly, **all models were tested with original raw data as it would be if EMS deployed a decision-support system in real life**. In this paper, the models were evaluated using the following regression metrics:

- Root mean squared error (RMSE) measures the square root average of the squares of the errors and is calculated as: $RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$;
- Mean absolute error (MAE) measures the averaged absolute difference between real and predicted values and is calculated as: $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$;
- Mean absolute percentage error (MAPE) measures how far the model's predictions are off from their corresponding outputs on average and is calculated as: $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\%$;
- Coefficient of determination (R^2) measures the proportion of the variance in the dependent variable that is predictable from the independent variable(s). An $R^2 = 1$ would indicate a model that fully captures the variation in ARTs;

in which y_i is the real output, $\hat{y}_i$ is the predicted output, and n is the total number of samples, for $i \in [1, n]$. Results for each metric were calculated considering the 6 months period of evaluation. The RMSE metric was also used during the hyperparameters tuning process via Bayesian optimization (BO). To this end, we used the HYPEROPT library [39] with 100 iterations for each model. Table 4 in Appendix A displays the range of each hyperparameter we considered in the BO, as well as the final configuration used to train and evaluate the models.

3. Results and Discussion

In this section, we present the results of our experimental validation (Subsection 3.1) and a general discussion (Subsection 3.2) including related work and limitations.

3.1. Privacy-preserving ART prediction

Figure 4 illustrates the impact of the level of GI for each ML model to predict ART according to each metric. As one can notice in this figure, for XGBoost, LGBM, and LASSO, there were minor differences between training models with original location data or sanitized ones. On the other hand, models trained with MLP performed poorly with GI-based data. In addition, by analyzing models trained with original data, while the smaller RMSE for LASSO is about 5.65, for more complex ML-based models, RMSE is less than 5.6, achieving 5.54 with XGBoost and LGBM. In comparison with the results of existing literature, lower R^2 scores and similar RMSE and MAE results were achieved in [40] to predict ART while using original location data only. With more details, Table 5 in Appendix A numerically exhibits the results from Figure 4.

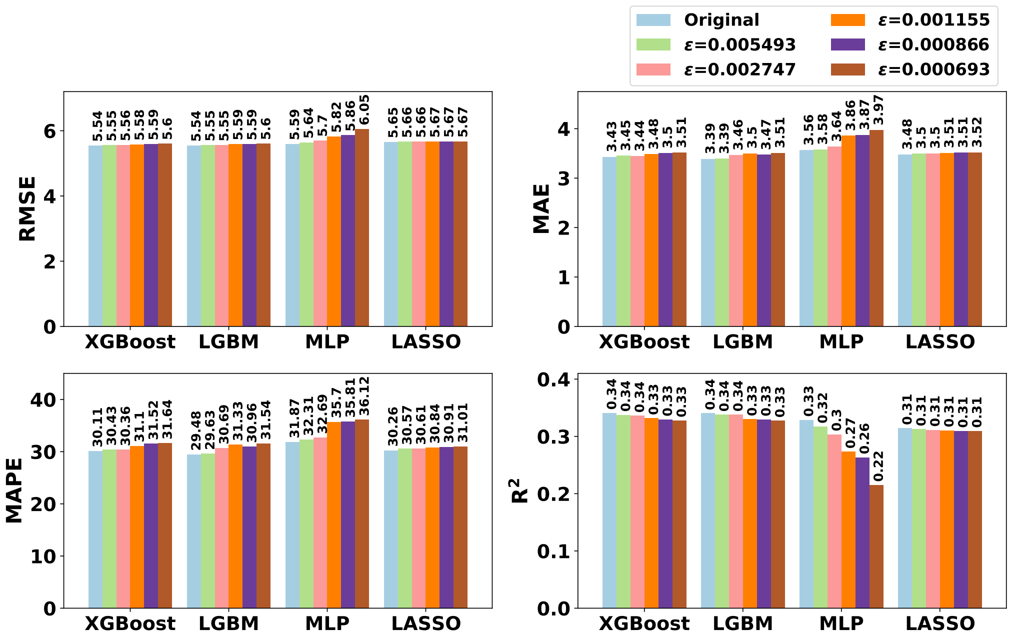

Figure 4. Impact of the level of ϵ-geo-indistinguishability for each ML model to predict ART according to each metric.

Indeed, among the four tested models, LGBM and XGBoost achieve similar metric results while favoring the LGBM model. Thus, Figure 5 illustrates the BO iterative process for LGBM models trained with original and sanitized data according to the RMSE metric (left-hand plot); and ART prediction results for 50 dispatched ambulances in 2020 out of 8,709 ones (right-hand plot) with an LGBM model trained with original data (Pred: original) and with two LGBM models trained sanitized data, *i.e.*, with $\epsilon = 0.005493$ (low privacy level) and with $\epsilon = 0.000693$ (high privacy level).

As one can notice in the left-hand plot of Figure 5, once data are sanitized with different levels of ϵ-GI, the hyperparameters optimization via BO is also perturbed. This way, local minimums were achieved in different steps of the BO (*i.e.*, the last marker per curve indicates the local minimum). For instance, even though $\epsilon = 0.002747$ is more strict than $\epsilon = 0.005493$, results were still better for the former since, in the last steps of BO, three better local minimums were found. Besides, prospective predictions were achieved with either original or sanitized data. For instance, in the right-hand plot of

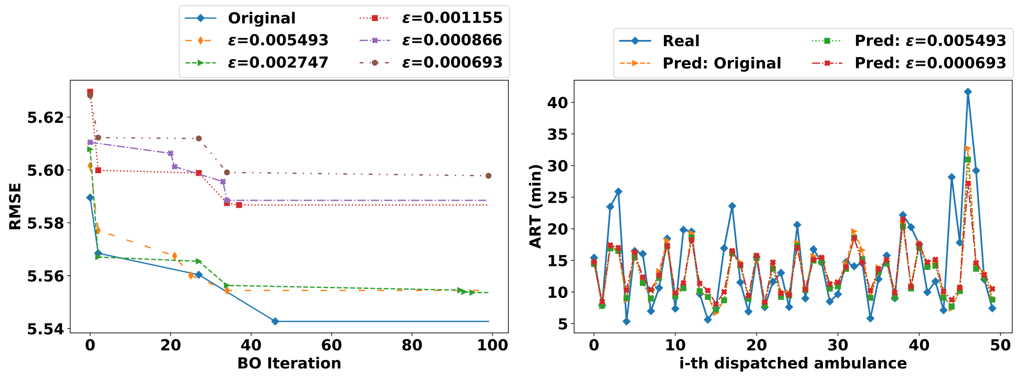

Figure 5. The left-hand plot illustrates the hyperparameters tuning process via Bayesian optimization with 100 iterations for LGBM models trained with original data and sanitized ones. The right-hand plot illustrates the prediction of ARTs with LGBM models trained with original data and with sanitized ones.

Figure 5, even for the high peak-value of ART around 40 minutes, LGBM's prediction achieved some reasonable estimation. Although several features were perturbed due to the sanitization of the emergency scene (*e.g.*, city, zone, etc), the models could still achieve similar predictions as the model trained with original location data.

Lastly, in general, the most important features considering LASSO coefficients and decision trees' importance scores were: OSRM API-based features (*i.e.*, estimated driving distance and estimated travel time); the great-circle distance between the center and the emergency scene; averaged ART per categorical features (*e.g.*, center, city, hour); the number of interventions in the previous hour, and the number of interventions still active.

3.2. Discussion

The medical literature has mainly focused attention on the analysis of ART [3,32,41] and its association with trauma [2,28] and cardiac arrest [1,4,6], for example. To reduce ART, some works propose reallocation of ambulances [5,42], operation demand forecasting [5,7,8,19,43], travel time prediction [11], simulation models [27,44], and EMS response time predictions [11,40]. The work in [40] propose a real-time system for predicting ARTs for the San Francisco fire department, which closely relates to this paper. The authors processed about 4.5 million EMS calls considering original raw location data to predict ART using four ML models, namely linear regression, linear regression with elastic net regularization, decision tree regression, and random forest. However, no privacy-preserving experiment was performed because the main objective of their paper was proposing a scalable, ML-based, and real-time system for predicting ART. Besides, we also included weather data that the authors in [40] did not consider in their system, which could help to recognize high ARTs due to bad weather conditions, for example.

Currently, many private and public organizations collect and analyze data about their associates, customers, and patients. Because most of these data are personal and confidential (*e.g.*, location), there is a need for privacy-preserving techniques for processing and using these data. Location privacy is an emergency research topic [12,13] due to the ubiquity of LBSs. Within our context, sharing and/or publishing the exact location of an emergency raises many privacy issues. For instance, the Seattle Fire Department [45] displays live EMS response information with the precise location and reason for the incident. While the intention of some fire departments [40,45] is laudable, there are many ways for (mis)using this information, which can jeopardize users' privacy. In our case, because the intervention's *reason* does not impose limits on SDIS 25 ARTs, we did not consider this sensitive attribute in our data analysis and privacy-preserving prediction models. Additionally, although during the EMS call processing, the SDIS 25 operator may acquire some personal data about the *victim*, this is not an operational requirement and, hence, we did not use this information too. This way, we focused our attention on the *location privacy* of each intervention.

To address location privacy, the authors in [14] proposed the concept of GI, which is based on a generalization [26] of the state-of-the-art DP [15] model. As highlighted in [14], attackers in LSBs may have side information about the user's reported location, *e.g.*, knowing that the user is probably visiting the Eiffel Tower instead of swimming in the Seine river. However, this does not apply in our context because someone may have drowned and EMS had to intervene. Similarly, even for the dataset with intermediate privacy (and higher) in which locations are spread out in the Doubs region (*cf.* map with 0.005493-GI location in Figure 3), someone may have been lost in the forest and EMS would have to interfere. For these reasons, sharing datasets with approximate emergency locations (*i.e.*, sanitized with GI, for example) has prospective directions as many locations are possible emergency scenes. Indeed, we are not interested in hiding the emergency's location completely since some approximate information is required in order to retrieve other features (*e.g.*, city, zone, estimated distance) to use for predicting ART.

Moreover, learning and extracting meaningful patterns from data, *e.g.*, through ML, play a key role in advancing and understanding several behaviors. However, on the one hand, storing and/or sharing raw personal data with trusted curators may still lead to data breaches [46] and/or misuse of data, which compromises users' privacy. On the other hand, training ML models with raw data can also leak private information. For instance, in [47] the authors evaluate how some models can memorize sensitive information from the training data, and in [48], the authors investigate how ML models are susceptible to membership inference attacks. To address these problems, some works [7,17–22,49] propose to train ML models with sanitized data, which is also known as input perturbation [23].

Input perturbation-based ML and GI are linked directly with local DP [23] in which each sample is sanitized independently, either by the user during the data collection process or by the trusted curator, which aims to preserve privacy of each data sample. This way, data are protected from data leakage and are more difficult to reconstruct, for example. In [20,49], the authors investigate how input perturbation through applying controlled Gaussian noise on data samples can guarantee (ϵ, δ)-DP on the final ML model. This means, since ML models are trained with perturbed data, there is a perturbation on the gradient and on the final parameters of the model too.

In this paper, rather than Gaussian noise, the emergency scenes were sanitized with Alg. 1, *i.e.*, adding two-dimensional Laplacian noise centered at the exact user location $x \in \mathbb{R}^2$. In addition, this sanitization also perturb other associated and calculated features such as: city, district, zone (*e.g.*, urban or not), great-circle distance, estimated driving distance, and estimated travel time (*cf.* Table 3). As well as the optimization of hyperparameters, *i.e.*, once data are differentially private, one can apply any function on it and, therefore, we also noticed perturbation on the BO procedure. Yet, as shown in the results, prospective ART predictions were achieved with either original or sanitized data. What is more, even with a high level of sanitization ($\epsilon = 0.000693$) there was a good privacy-utility trade-off. According to [50], if the mean absolute percentage error (*i.e.*, MAPE) is greater than 20% and less than 50%, the forecast is reasonable, which is the results we have in this paper with MAPE around 30%.

Lastly, some limitations of this work are described in the following. We analyzed ARTs using the data and operation procedures of only one EMS in France, namely SDIS 25. Although it may represent a sufficient amount of samples, other public and private organizations are also responsible for EMS calls, *e.g.*, the SAMU (Urgent Medical Aid Service in English) analyzed in [44]. Besides, there is the possibility of human error when using the mechanical system to report (*i.e.*, record) the arrival on-scene time "*ADate*". For instance, the crew may have forgotten to record status on arrival and may have registered later, or conversely, where the crew may have accidentally recorded before arriving at the location. Also, it is noteworthy to mention that the arrival on-scene does not mean arriving at the victim's side, *e.g.*, in some cases the real location of a victim is at the *n*-th stage of a building as investigated in [41].

4. Conclusion

In the event of an acute medical event such as a respiratory crisis or cardio-respiratory arrest, the time an ambulance takes to arrive on-scene has a direct impact on the quality of service provided [1,2,4–6,28]. Ambulance response time is a fundamental indicator of the effectiveness of EMS systems. For this reason, an intelligent decision-support system is necessary to help minimize overall EMS response times. The present work first analyzes historical records of ARTs to find correlations between their extracted features and explain the trends through the 15 years of collected data. Then, we sought to predict the response time that each center equipped with ambulances had to an event, but not only that, because we also consider that sharing or making public the location of the emergency would be subject to privacy issues. Therefore, the joint work aimed to evaluate the effectiveness of predicting ARTs considering ML models trained over sanitized location data with different levels of ϵ-geo-indistinguishability. As shown in the results, the sanitization of location data and the perturbation of its associated features (*e.g.*, city, distance) had no considerable impact on predicting ART. With these findings, EMS may prefer using and/or sharing sanitized datasets to avoid possible data leakages, membership inference attacks, or data reconstructions, for example.

For future work, we aim to extend the analysis and predictions to different operation times such as the pre-travel delay (*i.e.*, gathering personnel and ambulances) and travel time (*e.g.*, from the emergency scene to hospitals), while respecting users' privacy. In addition, new variables will be considered such as the number of dispatched ambulances registered in a previous or current time, and the number of ambulances and firefighters available in each center at a given time, given that while there are few resources available, ART may be longer. Indeed, the aim is to build an intelligent system capable of predicting ARTs while respecting victims' privacy. This way, this system would allow us to reinforce SDIS 25 centers with the necessary firefighters to attend incidents faster; to create a new center according to the concurrence and high average ARTs for a given area; as well as to convert a static resource deployment plan into a dynamic one, which would be based on the selection of the center with shorter response times taking into account the community the emergency took place, traffic and weather conditions, and so on.

Author Contributions: Conceptualization, H.H.A. and S.C.; methodology, H.H.A. and S.C.; software, H.H.A and S.C.; validation, H.H.A and S.C.; formal analysis, H.H.A and S.C.; investigation, H.H.A, S.C., C.G., and J.F.C.; resources, C.G. and J.F.C.; data curation, H.H.A and S.C.; writing—original draft preparation, H.H.A., S.C., C.G., and J.F.C.; writing—review and editing, H.H.A., S.C., C.G., and J.F.C.; visualization, S.C. and C.G.; supervision, C.G. and J.F.C.; project administration, C.G. and J.F.C.; funding acquisition, C.G. and J.F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: This work was supported by the Region of Bourgogne Franche-Comté CADRAN Project, by the EIPHI-BFC Graduate School (contract "ANR-17-EURE-0002"), and by SDIS du Doubs, with the support of the French Ministry of Higher Education and Research (managed by the National Association of Research and Technology (ANRT) for the CIFRE thesis (N 2019/0372). The authors would also like to thank SDIS 25 Commander Guillaume Royer-Fey and Captain Céline Chevallier for their great collaborations and continuous feedback. All computations have been performed on the "Mésocentre de Calcul de Franche-Comté".

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Ambulance response time	ART
Bayesian optimization	BO
Differential privacy	DP
Emergency medical services	EMS
Geo-Indistinguishability	GI
Least Absolute Shrinkage and Selection Operator	LASSO
Location-based services	LBSs
Local differential privacy	LDP
Light Gradient Boosted Machine	LGBM
Multilayer Perceptron	MLP
Mean absolute error	MAE
Mean absolute percentage error	MAPE
Root mean squared error	RMSE
Departmental Fire and Rescue Service of Doubs	SDIS 25
Extreme Gradient Boosting	XGBoost
Zone urban	Z1
Zone semi-urban	Z2
Zone rural	Z3

Appendix A. Complementary Results

Model	**Search space**	**Best configuration per dataset**					
		Original	$\epsilon = 0.005493$	$\epsilon = 0.002747$	$\epsilon = 0.001155$	$\epsilon = 0.000866$	$\epsilon = 0.000693$
XGBoost	max_depth: [1, 10]	9	9	6	6	9	9
	n_estimators: [50, 500]	465	465	130	235	465	465
	learning_rate: [0.001, 0.5]	0.0265	0.0265	0.0858	0.0486	0.0265	0.0265
	min_child_weight: [1, 10]	5	5	7	7	5	5
	max_delta_step: [1, 11]	4	4	3	4	4	4
	gamma: [0.5, 5]	3	3	0	2	3	3
	subsample: [0.5, 1]	0.8	0.8	1	1	0.8	0.8
	colsample_bytree: [0.5, 1]	0.5	0.5	0.5	0.5	0.5	0.5
	alpha: [0, 5]	2	2	1	2	2	2
LGBM	max_depth: [1, 10]	7	8	10	8	8	6
	n_estimators: [50, 500]	355	326	477	250	80	441
	learning_rate: [1e-4, 0.5]	0.0188	0.0098	0.0164	0.0285	0.0586	0.0300
	subsample: [0.5, 1]	0.54066	0.5228	0.6138	0.6699	0.6732	0.5812
	colsample_bytree: [0.5, 1]	0.5160	0.5575	0.5204	0.6870	0.5507	0.5451
	num_leaves: [31, 400]	400	192	245	398	132	95
	reg_alpha: [0, 5]	4	0	5	0	1	4
MLP	Dense layers: [1, 7]	7	3	4	6	6	6
	Number of neurons: $[2^8, 2^{13}]$	2^{10}	2^{12}	2^{12}	2^9	2^{12}	2^9
	Batch size: [32, 168]	140	80	48	82	70	44
	Learning rate: [1e-5, 0.01]	0.00265	0.00124	0.0099	0.0099	0.0094	0.0077
	Optimizer: Adam	Adam	Adam	Adam	Adam	Adam	Adam
	Epochs: 100	100	100	100	100	100	100
	Early stopping: 10	10	10	10	10	10	10
LASSO	alpha: [0.01, 2]	0.0205	0.0307	0.0105	0.0100	0.0112	0.0107

Table 4: Search space for hyperparameters by ML model and the best configuration obtained for predicting ARTs per dataset.

References

1. Bürger, A.; Wnent, J.; Bohn, A.; Jantzen, T.; Brenner, S.; Lefering, R.; Seewald, S.; Gräsner, J.T.; Fischer, M. The Effect of Ambulance Response Time on Survival Following Out-of-Hospital Cardiac Arrest. *Deutsches Aerzteblatt Online* **2018**. doi:10.3238/arztebl.2018.0541.
2. Byrne, J.P.; Mann, N.C.; Dai, M.; Mason, S.A.; Karanicolas, P.; Rizoli, S.; Nathens, A.B. Association Between Emergency Medical Service Response Time and Motor Vehicle Crash Mortality in the United States. *JAMA Surgery* **2019**, *154*, 286. doi:10.1001/jamasurg.2018.5097.
3. Do, Y.K.; Foo, K.; Ng, Y.Y.; Ong, M.E.H. A Quantile Regression Analysis of Ambulance Response Time. *Prehospital Emergency Care* **2012**, *17*, 170–176. doi:10.3109/10903127.2012.729127.

Data	Metric	XGBoost	LGBM	MLP	LASSO
Original	RMSE	**5.5398**	5.5427	5.5916	5.6511
	MAE	3.4286	**3.3880**	3.5623	3.4760
	MAPE	30.114	**29.476**	31.867	30.260
	R^2	**0.3412**	0.3405	0.3289	0.3145
$\epsilon = 0.005493$	RMSE	5.5547	**5.5544**	5.6401	5.6596
	MAE	3.4515	**3.3915**	3.5773	3.4960
	MAPE	30.432	**29.628**	32.307	30.571
	R^2	0.3377	**0.3378**	0.3172	0.3124
$\epsilon = 0.002747$	RMSE	5.5617	**5.5536**	5.6959	5.6636
	MAE	**3.4430**	3.4628	3.6357	3.4991
	MAPE	**30.364**	30.688	32.687	30.606
	R^2	0.3360	**0.3379**	0.3036	0.3115
$\epsilon = 0.001155$	RMSE	**5.5788**	5.5867	5.8184	5.6671
	MAE	**3.4803**	3.4991	3.8550	3.5094
	MAPE	31.097	31.327	35.704	**30.835**
	R^2	**0.3319**	0.3300	0.2733	0.3106
$\epsilon = 0.000866$	RMSE	5.5892	**5.5885**	5.8575	5.6716
	MAE	3.5033	**3.4702**	3.8736	3.5134
	MAPE	31.515	30.964	35.810	**30.907**
	R^2	0.3295	**0.3296**	0.2635	0.3095
$\epsilon = 0.000693$	RMSE	**5.5962**	5.5978	6.0463	5.6717
	MAE	3.5119	**3.5087**	3.9704	3.5171
	MAPE	31.638	31.543	36.122	**31.007**
	R^2	**0.3278**	0.3274	0.2153	0.3095

Table 5: Metrics results for each ML model trained with original data and sanitized ones. The best results per metric and model are highlighted in bold.

4. Holmén, J.; Herlitz, J.; Ricksten, S.E.; Strömsöe, A.; Hagberg, E.; Axelsson, C.; Rawshani, A. Shortening Ambulance Response Time Increases Survival in Out-of-Hospital Cardiac Arrest. *Journal of the American Heart Association* **2020**, *9*. doi:10.1161/jaha.120.017048.
5. Chen, A.Y.; Lu, T.Y.; Ma, M.H.M.; Sun, W.Z. Demand Forecast Using Data Analytics for the Preallocation of Ambulances. *IEEE Journal of Biomedical and Health Informatics* **2016**, *20*, 1178–1187. doi:10.1109/jbhi.2015.2443799.
6. Lee, D.W.; Moon, H.J.; Heo, N.H. Association between ambulance response time and neurologic outcome in patients with cardiac arrest. *The American Journal of Emergency Medicine* **2019**, *37*, 1999–2003. doi:10.1016/j.ajem.2019.02.021.
7. Arcolezi, H.H.; Couchot, J.F.; Cerna, S.; Guyeux, C.; Royer, G.; Bouna, B.A.; Xiao, X. Forecasting the number of firefighter interventions per region with local-differential-privacy-based data. *Computers & Security* **2020**, *96*, 101888. doi:10.1016/j.cose.2020.101888.
8. Cerna, S.; Guyeux, C.; Arcolezi, H.H.; Couturier, R.; Royer, G. A Comparison of LSTM and XGBoost for Predicting Firemen Interventions. In *Trends and Innovations in Information Systems and Technologies*; Springer International Publishing, 2020; pp. 424–434. doi:10.1007/978-3-030-45691-7_39.
9. Cerna, S.; Guyeux, C.; Royer, G.; Chevallier, C.; Plumerel, G. Predicting Fire Brigades Operational Breakdowns: A Real Case Study. *Mathematics* **2020**, *8*, 1383. doi: https://doi.org/10.3390/math8081383.
10. Nehme, Z.; Andrew, E.; Smith, K. Factors Influencing the Timeliness of Emergency Medical Service Response to Time Critical Emergencies. *Prehospital Emergency Care* **2016**, *20*, 783–791. doi:10.3109/10903127.2016.1164776.
11. Aladdini, K. EMS response time models: A case study and analysis for the region of Waterloo. Master's thesis, University of Waterloo, 2010.
12. Shokri, R.; Theodorakopoulos, G.; Boudec, J.Y.L.; Hubaux, J.P. Quantifying Location Privacy. 2011 IEEE Symposium on Security and Privacy. IEEE, 2011. doi:10.1109/sp.2011.18.
13. Chatzikokolakis, K.; ElSalamouny, E.; Palamidessi, C.; Pazii, A. Methods for Location Privacy: A comparative overview. *Foundations and Trends® in Privacy and Security* **2017**, *1*, 199–257. doi:10.1561/3300000017.
14. Andrés, M.E.; Bordenabe, N.E.; Chatzikokolakis, K.; Palamidessi, C. Geo-indistinguishability. Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security - CCS '13. ACM Press, 2013. doi:10.1145/2508859.2516735.
15. Dwork, C.; Roth, A.; others. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* **2014**, *9*, 211–407.
16. Location Guard. https://github.com/chatziko/location-guard.

17. Chamikara, M.A.P.; Bertok, P.; Khalil, I.; Liu, D.; Camtepe, S. Privacy Preserving Face Recognition Utilizing Differential Privacy. *Computers & Security* **2020**, *97*, 101951. doi: 10.1016/j.cose.2020.101951.
18. Fan, L. Image Pixelization with Differential Privacy. In *Data and Applications Security and Privacy XXXII*; Springer International Publishing, 2018; pp. 148–162. doi:10.1007/978-3-319-95729-6_10.
19. Couchot, J.F.; Guyeux, C.; Royer, G. Anonymously forecasting the number and nature of firefighting operations. Proceedings of the 23rd International Database Applications & Engineering Symposium on - IDEAS '19. ACM Press, 2019. doi:10.1145/3331076.3331085.
20. Fukuchi, K.; Tran, Q.K.; Sakuma, J. Differentially Private Empirical Risk Minimization with Input Perturbation. In *Discovery Science*; Springer International Publishing, 2017; pp. 82–90. doi:10.1007/978-3-319-67786-6_6.
21. Agrawal, D.; Aggarwal, C.C. On the design and quantification of privacy preserving data mining algorithms. Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '01. ACM Press, 2001. doi: 10.1145/375551.375602.
22. Agrawal, R.; Srikant, R. Privacy-preserving data mining. Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00. ACM Press, 2000. doi:10.1145/342009.335438.
23. Kasiviswanathan, S.P.; Lee, H.K.; Nissim, K.; Raskhodnikova, S.; Smith, A. What Can We Learn Privately? 2008 49th Annual IEEE Symposium on Foundations of Computer Science. IEEE, 2008. doi:10.1109/focs.2008.27.
24. Chaudhuri, K.; Monteleoni, C.; Sarwate, A.D. Differentially private empirical risk minimization. *Journal of Machine Learning Research* **2011**, *12*.
25. Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H.B.; Mironov, I.; Talwar, K.; Zhang, L. Deep Learning with Differential Privacy; Association for Computing Machinery: New York, NY, USA, 2016; CCS '16, p. 308–318. doi:10.1145/2976749.2978318.
26. Chatzikokolakis, K.; Andrés, M.E.; Bordenabe, N.E.; Palamidessi, C. Broadening the Scope of Differential Privacy Using Metrics. In *Privacy Enhancing Technologies*; Springer Berlin Heidelberg, 2013; pp. 82–102. doi:10.1007/978-3-642-39077-7_5.
27. Peleg, K.; Pliskin, J.S. A geographic information system simulation model of EMS: reducing ambulance response time. *The American Journal of Emergency Medicine* **2004**, *22*, 164–170. doi: 10.1016/j.ajem.2004.02.003.
28. Pons, P.T.; Markovchick, V.J. Eight minutes or less: does the ambulance response time guideline impact trauma patient outcome? *The Journal of Emergency Medicine* **2002**, *23*, 43–48. doi:10.1016/s0736-4679(02)00460-2.
29. Bison-Futé. Les Prévisions de Trafic. Available online: https://www.bison-fute.gouv.fr (accessed on 02 February 2021).
30. Météo-France. Données Publiques. Available online: https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32 (accessed on 02 February 2021).
31. Luxen, D.; Vetter, C. Real-time routing with OpenStreetMap data. Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems; ACM: New York, NY, USA, 2011; GIS '11, pp. 513–516. doi:10.1145/2093973.2094062.
32. Austin, P.C. Quantile Regression: A Statistical Tool for Out-of-hospital Research. *Academic Emergency Medicine* **2003**, *10*, 789–797. doi:10.1197/aemj.10.7.789.
33. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016. doi:10.1145/2939672.2939785.
34. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems 30*; Guyon, I.; Luxburg, U.V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds.; Curran Associates, Inc., 2017; pp. 3146–3154.
35. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep learning*; Vol. 1, MIT press Cambridge, 2016.
36. Chollet, F.; others. Keras. https://keras.io, 2015.
37. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **1996**, *58*, 267–288.
38. Pedregosa, F.; others. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

39. Bergstra, J.; Yamins, D.; Cox, D.D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. Proceedings of the 30th International Conference on International Conference on Machine Learning. JMLR, 2013, ICML'13, p. I–115–I–123.
40. Lian, X.; Melancon, S.; Presta, J.R.; Reevesman, A.; Spiering, B.; Woodbridge, D. Scalable Real-time Prediction and Analysis of San Francisco Fire Department Response Times. 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, 2019. doi:10.1109/smartworld-uic-atc-scalcom-iop-sci.2019.00154.
41. Silverman, R.A.; Galea, S.; Blaney, S.; Freese, J.; Prezant, D.J.; Park, R.; Pahk, R.; Caron, D.; Yoon, S.; Epstein, J.; Richmond, N.J. The "Vertical Response Time": Barriers to Ambulance Response in an Urban Area. *Academic Emergency Medicine* **2007**, *14*, 772–778. doi: 10.1111/j.1553-2712.2007.tb02350.x.
42. Carvalho, A.; Captivo, M.; Marques, I. Integrating the ambulance dispatching and relocation problems to maximize system's preparedness. *European Journal of Operational Research* **2020**, *283*, 1064–1080. doi:10.1016/j.ejor.2019.11.056.
43. Lin, A.X.; Ho, A.F.W.; Cheong, K.H.; Li, Z.; Cai, W.; Chee, M.L.; Ng, Y.Y.; Xiao, X.; Ong, M.E.H. Leveraging Machine Learning Techniques and Engineering of Multi-Nature Features for National Daily Regional Ambulance Demand Prediction. *International Journal of Environmental Research and Public Health* **2020**, *17*, 4179. doi:10.3390/ijerph17114179.
44. Aboueljinane, L.; Jemai, Z.; Sahin, E. Reducing ambulance response time using simulation: The case of Val-de-Marne department Emergency Medical service. Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC). IEEE, 2012. doi: 10.1109/wsc.2012.6465018.
45. Seattle Fire Department: Real-Time 911 Dispatch. Available online: http://www2.seattle.gov/fire/realtime911/ (accessed on 18 February 2021).
46. McCandless, D.; Evans, T.; Quick, M.; Hollowood, E.; Miles, C.; Hampson, D.; Geere, D. World's Biggest Data Breaches & Hacks, 2021. Available online: https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/ (accessed on 18 February 2021).
47. Song, C.; Ristenpart, T.; Shmatikov, V. Machine Learning Models that Remember Too Much. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017. doi:10.1145/3133956.3134077.
48. Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership Inference Attacks Against Machine Learning Models. 2017 IEEE Symposium on Security and Privacy (SP). IEEE, 2017. doi:10.1109/sp.2017.41.
49. Kang, Y.; Liu, Y.; Niu, B.; Tong, X.; Zhang, L.; Wang, W. Input Perturbation: A New Paradigm between Central and Local Differential Privacy. *arXiv preprint arXiv:2002.08570* **2020**.
50. Lewis, C. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*; Butterworth scientific, Butterworth Scientific, 1982.