# EXPHLOT: EXplainable Privacy Assessment for Human LOcation Trajectories

Francesca Naretto[1(✉)], Roberto Pellungrini[2], Salvatore Rinzivillo[3], and Daniele Fadda[3]

[1] University of Pisa, Pisa, Italy
francesca.naretto@di.unipi.it
[2] Scuola Normale Superiore, Pisa, Italy
roberto.pellungrini@sns.it
[3] ISTI CNR, Pisa, Italy
{salvatore.rinzivillo,daniele.fadda}@isti.cnr.it

**Abstract.** Human mobility data play a crucial role in understanding mobility patterns and developing analytical services across various domains such as urban planning, transportation, and public health. However, due to the sensitive nature of this data, accurately identifying privacy risks is essential before deciding to release it to the public. Recent work has proposed the use of machine learning models for predicting privacy risk on raw mobility trajectories and the use of SHAP for risk explanation. However, applying SHAP to mobility data results in explanations that are of limited use both for privacy experts and end-users. In this work, we present a novel version of the EXPERT privacy risk prediction and explanation framework specifically tailored for human mobility data. We leverage state-of-the-art algorithms in time series classification, as ROCKET and INCEPTIONTIME, to improve risk prediction while reducing computation time. Additionally, we address two key issues with SHAP explanation on mobility data: first, we devise an entropy-based mask to efficiently compute SHAP values for privacy risk in mobility data; second, we develop a module for interactive analysis and visualization of SHAP values over a map, empowering users with an intuitive understanding of SHAP values and privacy risk.

**Keywords:** Mobility Data · Privacy · Explainability

## 1 Introduction

The analysis of human mobility data is very important for the development of analytical services and for supporting decision-making processes in many sectors: urban planning [33], health [16] or tourism [7]. During the COVID-19 pandemic, for example, studying human mobility data helped understand and explain to the public how the infection spreads and propose good practices to stop it. Analyses


A. Bifet et al. (Eds.): DS 2023, LNAI 14276, pp. 325–340, 2023.
https://doi.org/10.1007/978-3-031-45275-8_22

in this field are usually conducted on large datasets containing information on the temporal sequences of locations visited by individuals, such as GPS tracks. This type of data, however, is very sensitive, as it can lead to the disclosure of personal information about an individual, such as the home location and place of work. For example, it has been proven that four spatiotemporal points may be sufficient to identify 95% of the individuals within a mobility dataset [20]. To address privacy risks associated with mobility data, various methodologies have been proposed to protect the privacy of the users, but they often involve modifying the data or Machine Learning (ML) models, compromising overall performance. Striking a balance between privacy protection and data quality requires reliable and efficient methods to quantify privacy risk. Pratesi *et al.* [25] proposed a risk assessment framework that computes privacy risk through the definition and simulation of various attack scenarios. While effective, this framework has drawbacks, including high time complexity and the need to re-compute the privacy risk for all data when new samples are added.

To mitigate these problems, Pellungrini et al. [22] proposed a ML approach for the computation of privacy risk based on individual and collective mobility features extracted from the data. Further improvements have been proposed by Naretto *et al.* [21] with the EXPERT framework, which implements a Long Short Term Memory neural network (LSTM) able to predict privacy risk directly from mobility data trajectories. In compliance with the EU General Data Protection Regulation, EXPERT also ensures the "right to explanation", proposing the use of SHAP (SHapley Additive exPlanations) [17], a well-known explainer based on SHAP values, which is commonly used for its stability and robustness. However, EXPERT has several limitations: *L1)* the LSTM training is time demanding and requires deep models to be effective; *L2)* SHAP can be efficiently applied only with specific heuristics tailored on specific ML models, like *DeepExplainer*, whereas general prediction models require a lot of time to be explained, since they rely on the combinatorial evaluation of the SHAP values; *L3)* the explanation provided by SHAP in the context of mobility data is not easy to interpret, given the high number of dimensions, and it gives limited information to non-technical users. Therefore, in this paper, we propose EXPHLOT, a framework tailored towards human mobility data that solves the aforementioned problems. To tackle *L1* we employ state-of-the-art ML models for sequential data (as INCEPTIONTIME [14] and ROCKET [8]) to speed up the training process. For *L2*, we propose a novel optimization heuristic based on entropy masks to execute efficiently SHAP permutation explainer for mobility data. For *L3*, we propose a visualization dashboard specifically tailored for the analysis of human mobility focused on both privacy risk and explanation, thus improving the fruition of the system for non-technical users.

The paper is structured as follows: in Sect. 2, we present the most relevant papers in the related literature; in Sect. 3 are reported the necessary definitions and notation; in Sect. 4 is presented our proposed framework; in Sect. 5 we show an application of our proposed framework to real human mobility data and provide an empirical evaluation.

## 2 Related Works

***Privacy Risk Assessment.*** In our work, we use the PRUDEnce framework from Pratesi *et al.* in [25], which allows for a systematic computation of privacy risk in a data-driven way. At its core, PRUDEnce is based on the principle of *k-anonymity* [29] as it computes privacy risk based on the size of the $k$-sets for each individual represented in the data. PRUDEnce has been extensively used in privacy risk assessment for a diverse range of data [23,24]. The high computational cost of PRUDEnce lead to the development of ML approaches that try to predict privacy risk instead of computing it. Pellungrini *et al.* [22] developed an approach based on Individual Mobility Profiles extracted from the data. Naretto et al. [21] proposed the EXPERT framework, which improves PRUDEnce in two ways: first, by developing a ML methodology able to predict risk directly from sequential data, secondly by explaining the privacy risk prediction using a set of methodologies like SHAP [17] and LIME [26]. Our EXPHLOT starts from the EXPERT and adds new improvements by integrating models and solutions that leverage domain-specific characteristics of mobility data. Several works are related to privacy risk assessment, mainly focused on applying classic risk assessment techniques to various privacy problems [32]. One of the most recent and relevant works in the field of privacy risk assessment is the work from Silva *et al.* [30], in which the authors provide an application of CRISP methodology and fuzzy logic to natural language processing tasks. Their work relies on the definition of a sensitivity level for the features possibly extracted from an individual's text and therefore is not entirely data-driven like our approach. For location-based data, Khalfoun *et al.* [15] proposed EDEN, a federated learning approach to location anonymization that is based on the FURIA federated learning framework for re-identification risk assessment. In their setting, they consider three types of attack: AP-Attack, POI-Attack, and PIT-Attack, considering spatial, temporal, and aggregated features. EDEN then selects the best privacy preservation technique with respect to this kind of assessment.

***Predictive Models for Human Mobility Data.*** In this section, we present the latest solution in the context of predictive models for human mobility data. EXPHLOT predicts the privacy risk directly on mobility data. For this task, one of the most applied ML models is the Long Short-Term Memory networks (LSTM) [13], a specific architecture belonging to Recurrent Neural Network (RNN), that are able to overcome some of the shortcomings of RNN, e.g., vanishing gradient in fully connected RNN. LSTM have been applied to human mobility data in many works [7,21,34]. Song *et al.* [31] use a LSTM network to develop a system for simulating and predicting human mobility and transportation model at a citywide level, while Altché *et al.* [1] use a LSTM to model vehicular movement on highways. LSTM have been also applied to predict the privacy risk in human mobility data [21]. However, the application of LSTM requires deep models to be effective and hence also a long training time. Recently, Fawaz *et al.* proposed INCEPTIONTIME [14], an ensemble of deep inception modules. This model achieves comparable performance as the LSTM reducing the learning time. Another recent proposal is ROCKET [8]. It is an ensemble method based on convo-

lutional kernels which transform the time series into features that are then used to train a linear classifier. This approach is very efficient and stable, allowing good generalization capabilities. In EXPHLOT we exploit both INCEPTIONTIME and ROCKET models to overcome the time limitation of EXPERT.

### 2.1 Explainability

Explainability is one of the most important modern lines of research in AI as it is crucial in achieving Trustworthy Artificial Intelligence. [3] provides a comprehensive overview of existing techniques for interpretability in ML, identifying two main types of explanation models: *global* and *local* explainers. Local explainers focus on explaining the results of predictions on single instance [11,18,27] while global explainers explain the logic of the whole machine learning model [5,6,10]. With EXPHLOT, we aim at explaining to the end user the reasons why he/she has a privacy risk exploiting local explanations for time series. In this context there are many recent methods, however, the majority of them are computationally inefficient and require a long training time [12]. In this work, we provide explanations by using SHAP [17], a well-known explainer based on SHAP values, which is commonly used for its stability and robustness of results.

## 3 Background

### 3.1 Privacy Risk Assessment Framework

In this paper, we consider each individual's mobility as a trajectory, i.e., a temporally ordered sequence of pairs, $T_u = (l_1, t_1), (l_2, t_2), \ldots, (l_m, t_m)$, where $l_i = \langle x_i, y_i \rangle$ is the location identified by the latitude $x_i$ and longitude $y_i$, while $t_i$ $(i = 1, \ldots, m)$ denotes the corresponding timestamp such that $\forall 1 \leq i \leq m$ $t_i < t_{i+1}$. We denote by $\mathcal{D} = T_1, \ldots, T_n$ the *mobility dataset* that describes the movements of $n$ individuals. In this paper, we simulate a privacy attack on human mobility data to acquire the ground truth to train our predictive model. Our attack is simulated using the PRUDEnce framework.

As mentioned in Sect. 2, PRUDEnce is based on *k-anonymity* [29], in which the privacy risk computation relies on the size of $k$-sets for each individual in the data. PRUDEnce has been utilized for privacy risk assessment in various data domains, such as purchase and mobility data [23,24]. The framework provides an effective approach to quantifying privacy risks and has demonstrated its applicability in diverse contexts. For these reasons, we have chosen the PRUDEnce methodology as the pre-processing step for computing privacy risk on raw mobility data in our work.

Technically, the privacy risk computation procedure of PRUDEnce is general and requires the definition of a privacy attack. The *privacy risk computation* defined in Prudence is the following:

1. Define an attack, based on a specific background knowledge category $B$;
2. Consider a set of background knowledge configurations $B_1, B_2, ..., B_m$;

3. For all the configurations $B_1, B_2, ..., B_m$, compute all the possible instances $b \in B_k$ and its probability of re-identification;
4. For each individual, select the maximum privacy risk, defined as the maximum probability of re-identification across all the instances $b \in B_k$.

Therefore PRUDEnce adopts an *exhaustive* privacy risk evaluation technique, by considering all the possible background knowledge the attacker could have over a given dataset (or dataview of the original dataset). For our purpose, we consider the case where each individual is represented by a single trajectory $T_u$ in $\mathcal{D}$. Formally, given a single individual $u$, the probability of re-identification is:

$$Pr_{\mathcal{D}}(T_u|b) = \frac{1}{\sum_{T_i \in \mathcal{D}}\{matching(T_i, b)\}} \tag{1}$$

where $\mathcal{D}$ is the dataset under analysis, $b$ the background knowledge instance considered and $T_u$ the trajectory under analysis. In essence, we compute the support for $b$ with respect to each trajectory in the dataset. The *matching* function formalizes how an adversary matches background knowledge $b$ to the data. $b$ is generated systematically, i.e., PRUDEnce performs exhaustive privacy risk assessment, among all possible $b \in B_k$. We simulate an attack where we assume that an adversary has access to some of the points in the trajectory of an individual, knowing a subsequence of the original trajectory with the relative order of the points.

Let $h$ be the number of locations $l_j$ of an individual $u$ known by the adversary and let $L(T_u)$ be the complete sequence of locations $l_j \in T_u$ visited by $u$ (i.e., regardless of time). The location sequence background knowledge is a set of configurations based on $h$ locations, defined as $B_h = L(T_u)^{[h]}$, where $L(T_u)^{[h]}$ denotes the set of all the possible $h$-subsequences of the elements in the set $L(T_u)$, i.e., each instance $b \in B_h$ is a subsequence of locations of length $h$. In each $b$, the order among the elements is preserved and known to the adversary. The *matching* function for this privacy attack is therefore defined as:

$$matching(T_i, b) = \begin{cases} 1, & \text{if } b \subseteq L(T_u) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Privacy Risk is the maximum probability of re-identification across all $b$:

$$Risk(u, \mathcal{D}) = max(Pr_{\mathcal{D}}(d = u|b)) \tag{3}$$

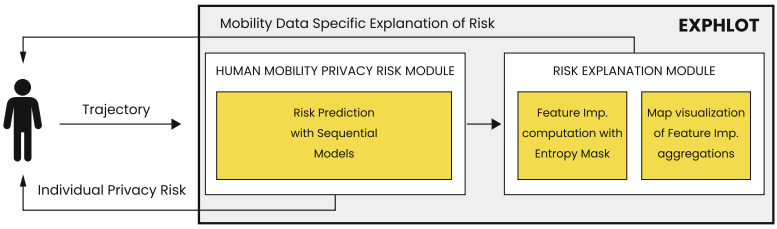


**Fig. 1.** The general structure of the proposed framework.

### 3.2 EXPERT

PRUDEnce is not suited for providing personalized recommendations in terms of risks associated with personal mobility: for any new user requiring risk evaluation, the system should re-compute the privacy risk against the whole dataset. In addition, it does not provide any explanation of the privacy risk derived by the system. To overcome these drawbacks, EXPERT [21] predicts the user's privacy risk to increase individual awareness, by also providing an explanation of the derivation of the risk associated with sharing sensitive location information. EXPERT implements a *privacy risk prediction* module which takes as input the user's trajectory and predicts the privacy risk level of that user by means of a ML model. It also uses an *explanation* module to produce the explanation of the predicted risk. The output of the *privacy risk prediction* module is the predicted privacy risk as a binary value (HIGH risk *vs* LOW risk). The output of the *risk explanation* module is an explanation of the ML model for the predicted risk label. EXPERT is modular with respect to the explainer, allowing the use of any explanation method which outputs a local explanation, suitable to the type of data under analysis. The authors use SHAP, and LORE in the original paper [21] (Fig. 1).

## 4 EXPHLOT

In this paper we propose EXPHLOT, an improved version of EXPERT tailored for Human Mobility Data. Our aim is to provide analysts with an actionable framework to predict and visualize privacy risk with an integrated explanation. The general architecture of EXPHLOT is shown in Fig. 1.

### 4.1 EXPHLOT Predictive Model

EXPHLOT objective is to predict the privacy risk of a human trajectory while providing the analyst with also an explanation to increase user awareness. Privacy risk is a continuous value in the interval $[0, 1]$. However, we decide to model the problem as a binary classification. Indeed, we are interested in distinguishing between HIGH risk and LOW risk users, in such a way that higher-risk users can be protected. Technically, we discretize the privacy risk obtained from the location-based attack: LOW risk or 0 (privacy risk $\leq 0.5$) and HIGH risk or 1 (privacy risk $>0.5$). The $\Gamma$ vector generated in this way is then joined to the mobility dataset $D$ and we use $\langle D, \Gamma \rangle$ to train a classification model. To avoid the problem of having to craft and compute features to be used as input data, Naretto *et al.* [21] propose to use methods applicable to raw sequences. In particular, they propose to solve the privacy risk classification problem using a Long-Short Term Memory network (LSTM). Our goal is, therefore, to use novel, state-of-the-art models to solve this prediction task, and to compare the performance and time-efficiency results of the new models with those of the LSTM. We propose two recent models, ROCKET and INCEPTIONTIME, introduced in Sect. 2. ROCKET is a fast and accurate time series classification algorithm that uses random convolutional kernels. It is composed of two parts: a first part in which $k$ randomly generated convolutional kernels are used to calculate a feature map from which, for each kernel,

two aggregated features are extracted (*ppv* and *maximum value*); a second part in which the aggregated features are passed to a linear classification algorithm to obtain the actual result. The number $k$ of kernels is the only hyper-parameter of the model. In theory, ROCKET can be used for both variable-length and fixed-length time series. To be applied to variable length time series, the kernels must be shorter than the length of the shortest time series. In the case where the length of the series varies greatly, as in our case, this approach is very inconvenient, as finding the right kernel would be time-consuming. We, therefore, chose a fixed-length approach, using low amplitude or zero padding to keep the result of the convolution operation on those segments close to zero and constant, cutting it off the calculation of the features (*ppv* and *maximum value*). We chose ROCKET over MINIROCKET [9] as the latter eliminates the random component in the choice of kernels' characteristics. Therefore, even though MINIROCKET is generally faster, we believe that a set of varied kernels fits better for our case, to capture the most diverse pattern possible. INCEPTIONTIME is an ensemble time series classification algorithm based on an ensemble of inception architectures. The Inception model is composed of convolutional layers and simultaneously applies several filters of different lengths to the input time series. This structure alleviates the vanishing gradient problem by enabling a direct flow of the gradient. It cannot be used on time series of variable length. To choose the best models, we focused on the recall of both classes, giving priority to class 1, and the precision of both classes. This is because we want to protect high-risk users by preventing them from being classified as LOW-risk, so that their sensitive data would not be threatened. Moreover, we wanted to maximize the possibility of sharing the data of LOW-risk users, thus preventing them from being classified as HIGH-risk.

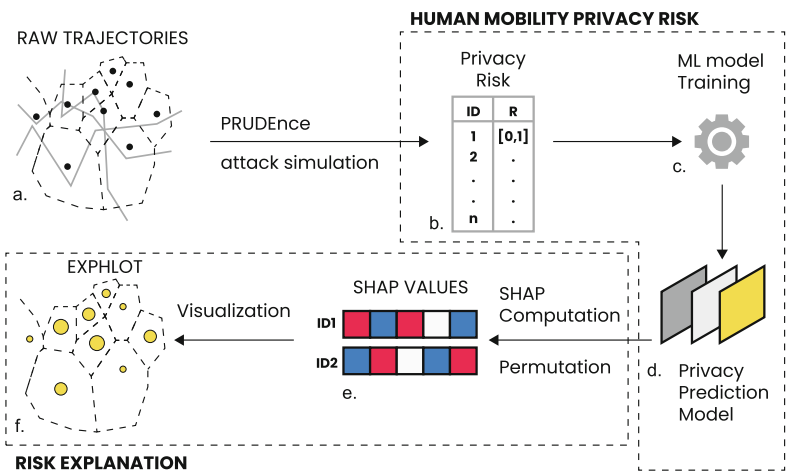


**Fig. 2.** EXPHLOT analytical pipeline. Starting from the generalized trajectories (a) a privacy prediction model (d) is trained from a set of observations generated by a privacy risk model (b). The prediction is explained by means of SHAP values (e) that are visualized within an analytical dashboard (f)

### 4.2 Exphlot Risk Explanation Module

For the Explanation Module of EXPHLOT our goal is to provide an explanation that is informative for experts and users in the dominion of Human Mobility data (Fig. 2). We chose to employ SHAP to generate an attribution-based explanation for our models. Our aim is to indicate, for each individual, what parts of his movement lead to higher privacy risks. Given the nature of our specific ML models, we must employ the *Kernel Explainer*, which is the agnostic explainer of the SHAP library. Clearly, depending on the size of the given data, the computation is more accurate but also longer in time. One possible solution, suggested also by the authors of SHAP, is to exploit K-means clustering by selecting a large $k$ and then feeding all the centroids obtained to the *Kernel Explainer*. In this way, we are able to represent all the space under analysis by considering a small number of trajectories. However, this solution for mobility data is not enough: SHAP considers each location of the trajectory as a variable and for computing the SHAP values all the permutations of variables are calculated as well as their relative interactions. This procedure is exponential in time if the number of variables is high, as in our case. Computation of SHAP values becomes therefore unfeasible in a reasonable time. Mitchell et al. [19] propose several sampling strategies that can in theory speed up SHAP values computation. However, many of the proposed strategies work under assumptions of bounds to the possible values or shape of the data. For human mobility, these bounds may not hold. For these reasons, we decide to apply the *PermutationExplainer* with a dynamic mask. This method can take as input a user-defined mask that allows certain features to be hidden, thus decreasing the individual evaluations made on these and the complexity of the calculation. In our setting, each feature corresponds to a location of the geographical map of our human mobility data. We used a binary mask to hide the features with the highest entropy, fully evaluating the locations with the lowest entropy. We formally define location entropy for each location $i$ in the dataset with the Shannon Entropy equation: $E_i = -\sum_{u \in U_i} p_u \log_2 p_u$, where $p_u$ is the probability that individual $u$ visits location $i$ and $U_i$ is the set of all individuals visiting location $i$. The importance of location entropy for privacy is thoroughly discussed by Rodriguez-Carrion *et al.* [28], while in the work of Pellungrini *et al.* [22] entropy is proven to be one of the most important predictive features/locations also in ML models. The intuitive concept behind it is that location entropy is a measure of anonymity, in the sense that if a user passes through high-entropy locations, where therefore many different other people pass through, the uniqueness of his mobility profile is lost as it is blurred by the general movement. We, therefore, hide the top 70% of the highest entropy locations, evaluating only the 30% with the lowest entropy. In this way, we are focusing on those locations that have fewer individuals visiting in a more sporadic way and thus we are focusing on explaining HIGH-risk predictions. Thus, we are able to speed up the computation of the SHAP values.

### 4.3 Exphlot Risk and Explanation Visualization Module

The effective visualization of mobility properties can provide a boost to gaining deeper insights into spatial and temporal patterns. To manage the complexity of spatial resolutions, a widely adopted solution leverages spatial aggregation based on spatial partitioning [2,4]. The process organizes close entities into groups and, for each group, a single centroid point is determined. Then the centroid points are used as seeds to partition the territory. In the scope of our work, the data related to geography is linked to multiple dimensions and attributes, like mobility indicators, privacy risk prediction, and feature relevance. Moreover, many of these indicators may have multiple spatial scales, for example ranging from an urban building block resolution to a city district.

Thus, we designed a visual interface where the set of locations of each trajectory is presented within two linked displays: a *dynamic map* with embedded graphics and a *bubble chart* (see Fig. 4). The *dynamic map* shows for each location a visual mark, a circle, whose visual properties are linked to internal indicators of the location it represents. Each circle is driven by two visual variables, the area of the circle and the fill color, which both encode the same quantitative value. Without loss of generality, we can assume that these quantitative values are mapped to the $[0, 1]$ interval, in order to implement a pair of scale functions to determine the area and the color of each circle. The *Bubble Chart* contains the same set of circles of the map (to create conceptual links between the two displays) located accordingly to the respective values on the two axes. The user can decide which attributes are associated with which value. Any selection/filter activated on the Bubble Chart is propagated to the map (and *viceversa*).

The SHAP values are computed for every single individual trajectory. However, the domain expert is interested in the analysis of collective behavior. Thus, we aggregate the individual explanations into a global one using the aggregation procedure available within the SHAP library. This is especially important for all those instances where the data is not public or is under strict confidentiality constraints. From a geographical point of view, we considered for each location $l$ the set of all the trajectories crossing $l$. For this subset of trajectories, a set of indicators is computed, such as *number of trajectories*, and *risk of re-identification*. For the latter, we compute statistical indicators to have a compact representation of the distribution: min, max, first quartile, third quartile, median, and average.

This design achieves multiple objectives. First, it provides a user-driven exploration of the SHAP values, since the analyst can evaluate and compare the contribution of each location to the risk prediction and let the user visually identify zones containing locations with similar characteristics. Second, the possibility of navigating the map allows for a deeper investigation of local areas and provides a solution to limit cluttering when the number of locations is high. Third, geographic mapping allows a topological exploration of close locations, enabling the identification of general patterns, i.e. urban areas versus rural areas. Fourth, the expert can exploit the linked display to investigate relevant cases that are not directly evident from the map. The possibility of cross-selecting visual elements enables better identification of patterns and rules of the data.

## 5 Experiments

For validating EXPHLOT we used GPS tracks of private vehicles, provided by Octo Telematics[1], an insurance company. We selected trajectories from the city area of Prato and Pistoia (Italy), with 8651 users observed in a period of one month, from 1st May to 31st May 2011[2]. The dataset considered is composed of a trajectory for each user. Hence, each trajectory contains all the points visited by the user in temporal order. On these trajectories, we applied a transformation, in the following called VORONOI, in which the territory is split in tiles based on a data-driven Voronoi tessellation [2]. This approach considers the traffic density of an area to create the tiles. Then, we used the cells of this tessellation to generalize the original trajectories. The algorithm applies interpolation between non-adjacent points[3]. The outliers were removed using DBScan algorithm obtaining 1473 different locations, with an average length of 240.2 per trajectory. Given the processed dataset $D$, for an in-depth validation of EXPHLOT, we considered *four background knowledge* configurations $B_h$ using $h = 2, 3, 4, 5$ obtaining four different risk datasets, $\Gamma_{h=2,3,4,5}$ where, we recall, $h$ represents the length of the background knowledge of the simulated attacker. We discretized the risk values in two classes: LOW, when the privacy risk is in $[0, 0.5]$ and HIGH in $]0.5, 1]$. At this point, we merged the privacy risk data with the trajectories to obtain the classification datasets for our supervised learning task, following the methodology explained in Sect. 3.1. Hence, we obtained 4 different datasets for our experiments. We remark that the datasets with the highest and lowest background knowledge are highly imbalanced, having the $D_{h=2}$ with the 71% of users belonging to the LOW class, while for $D_{h=5}$ has the 63% of trajectories in the HIGH class. This is to be expected, as when the knowledge of the attacker is small, such as $h = 2$, the attack is less effective, having fewer people re-identified. In addition, we remark that we compute the privacy risk of the entire dataset $D$, splitting the data after privacy risk computation. This decision is based on the fact that if we calculate the privacy risk separately for the training and testing sets, the final result will differ from the computation performed on the complete dataset, due to *k-anonymity* (Sect. 3.1). It has been demonstrated that the models still generalize well and possess transfer learning capabilities [22].

### 5.1 Exphlot Privacy Risk Prediction Module

For all the models we split our datasets into 80% for training and validation (10%) and 20% for testing. The predictive performance of ROCKET, INCEPTIONTIME, and LSTM are reported in Table 1. All the models perform well, achieving good precision and recall for both classes, even in unbalanced settings. For the most unbalanced case, which is the $h = 2$, ROCKET and INCEPTIONTIME

[1] https://www.octotelematics.com/it/.

[2] Data are collected by GPS devices that detect the position every 30 s, if the vehicle is not in motion the device automatically stops recording.

[3] Voronoi tessellation obtained using http://geoanalytics.net/V-Analytics.

**Table 1.** Metrics of ROCKET (R), INCEPTIONTIME (IT) and LSTM (LS) compared for each dataset $h$. For precision $P$ and recall $R$ we present the values for both classes (*high* and *low* risk. From a privacy perspective $R_{high}$ is the most important value as it represents the fraction of correctly predicted HIGH risk individuals.

| | $h=2$ | | | $h=3$ | | |
|---|---|---|---|---|---|---|
| | ROCKET | INCEPTIONTIME | LSTM | ROCKET | INCEPTIONTIME | LSTM |
| Acc | 0.81 | **0.84** | 0.80 | **0.88** | 0.87 | **0.88** |
| $P_{low}$ | **0.91** | 0.88 | 0.90 | 0.89 | 0.86 | **0.90** |
| $P_{high}$ | 0.63 | **0.72** | 0.62 | **0.88** | **0.88** | **0.88** |
| $R_{low}$ | 0.81 | **0.89** | 0.81 | 0.84 | **0.85** | 0.84 |
| $R_{high}$ | **0.80** | 0.70 | 0.76 | 0.91 | 0.89 | **0.92** |
| F1 | 0.78 | **0.80** | 0.76 | **0.88** | 0.87 | **0.88** |
| | $h=4$ | | | $h=5$ | | |
| | ROCKET | INCEPTIONTIME | LSTM | ROCKET | INCEPTIONTIME | LSTM |
| Acc | **0.90** | 0.89 | 0.89 | 0.91 | 0.90 | **0.92** |
| $P_{low}$ | **0.90** | 0.87 | **0.90** | 0.87 | 0.86 | **0.89** |
| $P_{high}$ | 0.90 | **0.91** | 0.89 | 0.93 | 0.93 | **0.94** |
| $R_{low}$ | 0.86 | **0.89** | 0.84 | 0.88 | 0.88 | **0.89** |
| $R_{high}$ | **0.93** | 0.90 | 0.92 | 0.92 | 0.92 | **0.93** |
| F1 | **0.90** | 0.89 | 0.89 | 0.90 | 0.90 | **0.91** |

**Table 2.** Training and test times for ROCKET and InceptionTime. Overall ROCKET is the fastest model in training.

| | INCEPTIONTIME | | ROCKET | | LSTM | |
|---|---|---|---|---|---|---|
| Dataset | Training | Test Time | Training | Test | Training | Test |
| $h=2$ | 16h49min | 6sec | 2min32sec | 44sec | 8h50min | 60sec |
| $h=3$ | 20h7min | 6sec | 3min | 40sec | 5h30min | 60sec |
| $h=4$ | 4h | 4sec | 7min | 16sec | 5h50min | 60sec |
| $h=5$ | 9h24min | 5sec | 8min | 17sec | 6h15min | 60sec |

perform better than LSTM, showing better generalization capabilities. However, ROCKET achieves the highest recall on class HIGH, which is the most important class for our setting, being the class of the users with HIGH risk of privacy. INCEPTIONTIME, instead, while having generally good metrics, does not perform well on the recall for HIGH class. The real benefit of ROCKET over other models is in training time, as can be seen in Table 2. While training the LSTM can take many hours, the other models are faster. ROCKET is the quickest, with a training time of just a few minutes, allowing us to achieve the *online* interaction with the end user we are aiming at.

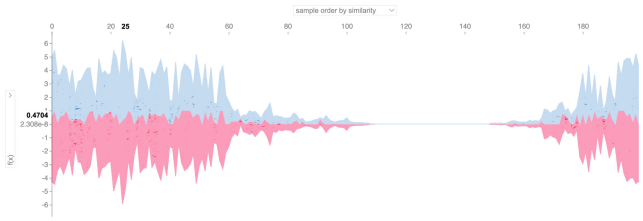


**Fig. 3.** Shap Force Plot visualization of the contributions towards HIGH risk. The standard visualization does not provide significant information to domain experts.

### 5.2 Mobility Privacy Risk Explanation

Applying SHAP we obtain a local explanation based on feature importance: for each feature we have a value associated to it that represents how important the feature is for the prediction at hand. Local explanations can be summed up to obtain a global explanation as shown in Fig. 3. This plot represents the explanation for all trajectories predicted as HIGH risk. A large number of features makes it very difficult for the analyst to understand which are the most relevant locations that contribute to the HIGH (or LOW) risk. Clearly, this linear layout has two main limitations: first, the high number of features does not allow a clear reading of those locations with smaller contributions; second, the topological and spatial relations among locations are not evident. The visual interface introduced in Sect. 4.3 addresses these two limitations. Figure 4 shows a screenshot of the interface showing the SHAP values associated with the prediction of HIGH risk for each location[4]. This visualization allows an analyst to immediately understand which areas of the map present the highest contribution for the model towards risk classification. Our map allows for a more intuitive understanding of the contributions of each location with respect to the original SHAP visualization. Our visualization can help the analyst understand the dependence of privacy risk on the mobility behaviors of the collectivity. In the figure, there is a cluster of locations along a country road with a high contribution to the HIGH risk, confirming the intuition that low-traffic roads are more prone to privacy exposures. Moreover, the urban surroundings present a lower level of risk, even if it is possible to visually detect different privacy levels in two close municipalities: the south-east town has very low-risk levels; the north-west town has a higher risk level.

[4] The interactive maps of the experiments in this paper are available at this link.

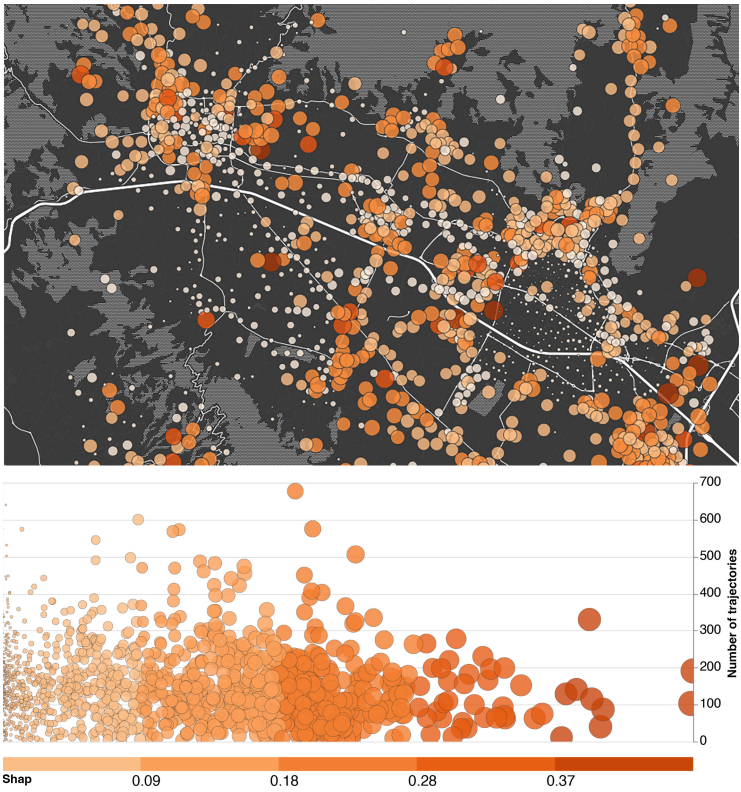


**Fig. 4.** Visual interface for the exploration of explanation and prediction of privacy risk. Each circle represents the contribution to the prediction of HIGH risk

## 6 Conclusion

In this paper, we proposed EXPHLOT, a privacy assessment prediction and explanation framework tailored towards human mobility data. We improve on previous privacy risk assessment frameworks by employing specific ML models for sequential data and develop custom heuristic techniques for computing SHAP values in feasible times and a visualization tool tailored for human mobility data analysis. Our framework can accurately predict privacy risk in human mobility data and effectively explain the predictive models with fast SHAP value calculation and an intuitive and interactive visualization tool that maps the essential contribution and information about the problem onto a dynamic map. We validated our framework on real, confidential human mobility data and showed how it is possible to immediately gain new insight into the nature of privacy risk. Our work provides privacy analysts and experts in the field with an interactive and actionable tool to understand the privacy risk of human mobility data in an interactive and fast way. As a future work, we are working on exploiting our visual analytics environment to validate the effeect of different privacy mitigation techniques. This would be a *"what-if"* simulation module to allow analysts

to interactively assess privacy risk, providing a new tool in the development of privacy protection measures based on generalization or deletion. Another interesting direction is the integration of additional data quality measures, to allow further experimentation of protection measures on the data before release.

**Acknowledgments.** This work is supported by: the EU NextGenerationEU programme under the funding schemes PNRR-PE-AI FAIR (Future Artificial Intelligence Research); the EU - Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 - Integrating Activities for Advanced Communities" (G.A. n.871042) "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu); PNRR-"SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics" - Prot. IR0000013; TAILOR (G.A. 952215), HumanE-AI-Net (G.A. 952026) and EU H2020 project XAI (Grant Id 834756); CREXDATA (G.A. 101092749).

## References

1. Altché, F., de La Fortelle, A.: An LSTM network for highway trajectory prediction. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), pp. 353–359 (2017)
2. Andrienko, N.V., Andrienko, G.L.: Spatial generalization and aggregation of massive movement data. IEEE Trans. Vis. Comput. Graph. **17**(2), 205–219 (2011)
3. Bodria, F., Giannotti, F., Guidotti, R., Naretto, F., Pedreschi, D., Rinzivillo, S.: Benchmarking and survey of explanation methods for black box models. DAMI (2023)
4. Buchmüller, J., Janetzko, H., Andrienko, G.L., Andrienko, N.V., Fuchs, G., Keim, D.A.: Visual analytics for exploring local impact of air traffic. Comput. Graph. Forum **34**(3), 181–190 (2015). https://doi.org/10.1111/cgf.12630
5. Craven, M., Shavlik, J.W.: Extracting tree-structured representations of trained networks. In: NIPS, pp. 24–30 (1996)
6. Craven, M.W., Shavlik, J.W.: Using sampling and queries to extract rules from trained neural networks. In: JMLR, pp. 37–45. Elsevier (1994)
7. Crivellari, A., Beinat, E.: LSTM-based deep learning model for predicting individual mobility traces of short-term foreign tourists. Sustainability **12**(1) (2020). https://doi.org/10.3390/su12010349
8. Dempster, A., Petitjean, F., Webb, G.I.: ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. Data Min. Knowl. Disc. **34**(5), 1454–1495 (2020). https://doi.org/10.1007/s10618-020-00701-z
9. Dempster, A., Schmidt, D.F., Webb, G.I.: MiniRocket. In: Proceedings of the 27th ACM SIGKDD Conference. ACM (2021). https://doi.org/10.1145/3447548.3467231
10. Deng, H.: Interpreting tree ensembles with inTrees. Int. J. Data Sci. Anal. **7**(4), 277–287 (2019)
11. Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., Turini, F.: Factual and counterfactual explanations for black box decision making. IEEE Intell. Syst. **34**(6), 14–23 (2019)
12. Guidotti, R., Monreale, A., Spinnato, F., Pedreschi, D., Giannotti, F.: Explaining any time series classifier. In: CogMI 2020 (2020)
13. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

14. Ismail Fawaz, H., et al.: InceptionTime: finding alexnet for time series classification. Data Min. Knowl. Discov. **34**, 1936–1962 (2020)
15. Khalfoun, B., Ben Mokhtar, S., Bouchenak, S., Nitu, V.: Eden: Enforcing location privacy through re-identification risk assessment: a federated learning approach (2021). https://doi.org/10.1145/3463502
16. Lucchini, L., et al.: Living in a pandemic: changes in mobility routines, social activity and adherence to COVID-19 protective measures. Sci. Rep. (2021). https://doi.org/10.1038/s41598-021-04139-1
17. Lundberg, S.M., Lee, S.: A unified approach to interpreting model predictions. CoRR abs/1705.07874 (2017). http://arxiv.org/abs/1705.07874
18. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: NIPS, pp. 4765–4774 (2017)
19. Mitchell, R., Cooper, J., Frank, E., Holmes, G.: Sampling permutations for shapley value estimation. J. Mach. Learn. Res. **23**, 1–46 (2022)
20. Montjoye, Y.A., Hidalgo, C., Verleysen, M., Blondel, V.: Unique in the crowd: the privacy bounds of human mobility. Sci. Rep. (2013). https://doi.org/10.1038/srep01376
21. Naretto, F., Pellungrini, R., Nardini, F.M., Giannotti, F.: Prediction and explanation of privacy risk on mobility data with neural networks. In: ECML PKDD 2020 Workshops (2020)
22. Pappalardo, L., Pellungrini, R., Pratesi, F., Monreale, A.: A data mining approach to assess privacy risk in human mobility data. ACM Trans. Intell. Syst. Technol. (2017). https://doi.org/10.1145/3106774
23. Pellungrini, R., Pappalardo, L., Pratesi, F., Monreale, A.: Analyzing privacy risk in human mobility data (2018)
24. Pellungrini, R., Pratesi, F., Pappalardo, L.: Assessing privacy risk in retail data (2017)
25. Pratesi, F., Monreale, A., Trasarti, R., Giannotti, F., Pedreschi, D., Yanagihara, T.: Prudence: a system for assessing privacy risk vs utility in data sharing ecosystems. Trans. Data Priv. **11**, 139–167 (2018)
26. Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you?: explaining the predictions of any classifier (2016)
27. Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you?: explaining the predictions of any classifier. In: ACM SIGKDD, pp. 1135–1144 (2016)
28. Rodriguez-Carrion, A., et al.: Entropy-based privacy against profiling of user mobility. Entropy **17**(6), 3913–3946 (2015). https://doi.org/10.3390/e17063913
29. Samarati, P.: Protecting respondents identities in microdata release. IEEE Trans. Knowl. Data Eng. (2001). https://doi.org/10.1109/69.971193
30. Silva, P., Gonçalves, C., Antunes, N., Curado, M., Walek, B.: Privacy risk assessment and privacy-preserving data monitoring. Expert Syst. Appl. **200** (2022)
31. Song, X., Kanasugi, H., Shibasaki, R.: Deeptransport: prediction and simulation of human mobility and transportation mode at a citywide level. In: IJCAI'16 (2016)
32. Tang, J., Cui, Y., Li, Q., Ren, K., Liu, J., Buyya, R.: Ensuring security and privacy preservation for cloud data services. ACM Comput. Surv. (CSUR) **49**, 1–39 (2016)
33. Wang, J., Kong, X., Xia, F., Sun, L.: Urban human mobility: data-driven modeling and prediction. SIGKDD Explor. Newsl. **21**, 1–19 (2019)
34. Wu, F., Fu, K., Wang, Y., Xiao, Z., Fu, X.: A spatial-temporal-semantic neural network algorithm for location prediction on moving objects. Algorithms (2017). https://doi.org/10.3390/a10020037