# Exploiting Points of Interest for Predictive Policing

Luís Gustavo Coutinho do Rêgo
gustavo.coutinho@insightlab.ufc.br
Insight Data Science Lab
Brazil

Ticiana Linhares Coelho da Silva
ticianalc@insightlab.ufc.br
Insight Data Science Lab
Brazil

Regis Pires Magalhães
regis@insightlab.ufc.br
Insight Data Science Lab
Brazil

José Antônio Fernandes de Macêdo
jose.macedo@insightlab.ufc.br
Insight Data Science Lab
Brazil

Wellington Clay Porcino Silva
wellington.porcino@mj.gov.br
National Department of Public Security
Brazil

## ABSTRACT

High crime rates have become a public health problem in many important cities, according to World Health Organization. Many researchers have been developing algorithms to predict crime occurrences to tackle this problem. The smart cities' environment can provide us enough ubiquitous data, e.g., traffic flow, human mobility, and Points of Interest (POI) information, to feed those predictive policing algorithms and reflect city dynamics. POIs data provide essential information such as geographical location, category, customer reviews, and busy hours. Recent studies have shown that POI geographical locations are useful for predictive policing. In this paper, we aim at predicting crimes in a delimited region around the POIs of a city with new environmental features. We investigate the relevance of POIs location and the semantic and the temporal features from POIs data in our problem. We also propose and analyze different machine learning approaches to train prediction functions based on these features and conduct experiments on real crime data over multiple years. The experiments demonstrate that the popular time feature is more relevant than the historical information about the number of crimes around a POI, but both information is much less critical than the spatio-temporal information. This work is the first that studies the popular time feature extracted from POIs data and historical criminal information for predictive policing from the authors' knowledge.

## CCS CONCEPTS

• **Information systems → Geographic information systems**.

## KEYWORDS

predictive policing, point of interest, spatial-temporal systems


*ARIC'20, November 3–6, 2020, Seattle, WA, USA*

ACM ISBN 978-1-4503-8165-9/20/11…$15.00
https://doi.org/10.1145/3423455.3430319

**ACM Reference Format:**
Luís Gustavo Coutinho do Rêgo, Ticiana Linhares Coelho da Silva, Regis Pires Magalhães, José Antônio Fernandes de Macêdo, and Wellington Clay Porcino Silva. 2020. Exploiting Points of Interest for Predictive Policing. In *3rd ACM SIGSPATIAL Workshop on Advances in Resilient and Intelligent Cities (ARIC'20), November 3–6, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3423455.3430319

## 1 INTRODUCTION

Crimes have emerged as one of the most critical problems countries face, making the prevention of violence a public health priority [13]. In particular, in Brazil, crime is the prime concern in some cities due to the high crime rates, the sheer magnitude of violence, and the perceived number of lives lost. In 2018, in Fortaleza City, the 5th most populous in Brazil, citizens reported approximately 53,000 robberies, more than 4,000 homicides, and over 10,000 thefts according to police reports [8]. In response to those numbers, public police departments could use intelligent police tactics and improvements in law-enforcement analytics to handle the high number of criminal occurrences [17]. As an example of smart tactics that could be adopted by the police, we have the use of techniques to visualize criminal stains to help in the allocation of police patrols, characterization of criminal environments based on occurrences and environmental data, and the use of machine learning models in support decision making.

Therefore, to improve citizens' life quality, accurate and reliable prediction of crimes is a necessity for helping governments and police departments effectively prevent crimes from happening and handle them efficiently when they occur [12].

Many algorithms used for crime predictions rely on a large amount of data for model training. With the technological growth, smart cities can provide us enough ubiquitous data, e.g., traffic flow, human mobility, and Points of Interest (POI) information, to reflect city dynamics and provide new insights to understand some public security problems [3], such as the predictive policing problem addressed in this paper. In other words, smart cities' advancement can better describe the environments in which citizens live with all their specific needs, particularly the needs related to public security. POIs data provides information such as the GPS coordinates, category, customer reviews, and popular time, which shows how busy a POI typically is during different times. Recent studies have shown that POIs locations are useful for predictive policing [12, 19, 24] and crime rate analysis [20]. This information makes sense since entertainment, business, and sustenance facilities may have higher crime rates in the area around them [2]. For instance, robberies usually occur in places nearby shops, bars, and restaurants, where people often move around.

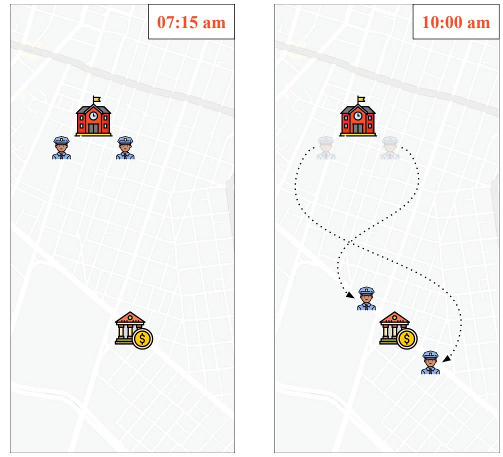
In this paper, we address the problem of predicting the number of crime incidents for a 100m circle region around a POI and the influence of spatial-temporal and semantic features. First, we study the spatial factors of crime, analyzing the geographical impact by considering the POIs location (latitude and longitude). For simplicity, we used a 100m circle region due to the strong influence of the Points of Interest on street robbery count in the block level, as shown in [2]. Second, the semantic feature considered is the popular time captured for each POI. This feature shows how busy a POI is during different time slots. Popularity for any given time slot is shown relative to the POIs typical peak popularity in a week. We believe this information is relevant since crime is more likely to happen when people visit a POI.

As an example of how the semantic information about the popular times of POIs could be used, Figure 1 shows how police units could be allocated. At 7:15 am, police units should be patrolling around schools since this is when children arrive at this category of POI, i.e., the period where there is a high value for the popular time feature (Figure 1a). Then, at 10:00 am, police units should be reallocated to, for example, POIs with bank category (Figure 1b). In this new time, the popular time feature of the POIs with the school category should have decreased, and police action in that area is no longer necessary. On the other hand, the popular time feature value for the POIs with the bank category should have increased, demanding police patrol.

Finally, we also explore temporal features since factors underlying crime occurrences may change over time. For instance, burglary causality in the morning may differ from the night in urban areas, and crime causality on weekdays may differ from weekends. Moreover, as stated in [7], there is a temporal patterning of crime incidents. The occurrence of a crime actively increases the probability of further incidents in the vicinity [7].

Crime prediction based on crime occurrences and POIs location and popular times could be used by police departments to allocate, more intelligently and strategically, police officers in areas with more significant citizens' movement, such as near bars, restaurants, bus stops, and so avoid new crime occurrences.

We address our problem by proposing and analyzing different machine-learning approaches to train different prediction functions. In summary, this paper's contributions are: 1) we exploit different features (spatial, temporal, and semantic) from POI and analyze their relevance for the crime prediction problem. We also combine these features from POIs data with historical crime occurrences for learning predictive policing functions; 2) We conduct extensive experiments, including comparing different machine learning prediction models. We conducted experiments on real crime data from 2014 to 2019 in Fortaleza, Ceará, Brazil. The experiments demonstrate that the spatio-temporal information from POIs provides way more relevant features for predictive policing. Nevertheless, the experiments also demonstrated that the popular time feature is more relevant than the historical information about the number of crimes around a POI. From the best of the authors' knowledge, this is the first work that studies different features extracted from POIs data and their predictive policing relevance.




**(a) Initial police units position.** **(b) Police units movement after a period.**

**Figure 1: Example of how the police units allocation could change depending on the popular time of a Point of Interest.**

The remainder of the paper is structured as follows: Section 2 formally defines the problem. Section 3 presents the related works. Section 4 discusses we prepare the dataset to build the prediction models. Section 5 discusses the experimental evaluation. And finally, Section 6 draws the final conclusions.

## 2 PRELIMINARIES

This section will first perform a preliminary analysis of the correlation between POIs and crime rates. Next, we will introduce some necessary notations and formally present the predictive policing problem formulation.

### 2.1 An Analysis of Correlation Between POIs and Crime Rate

Let a road network be represented by a graph $G(V, E)$, where $V$ represents the set of nodes and $E$ the set of edges (road segments). A POI $p$ represents an object within the underlying network characterized by its geographic coordinates (latitude and longitude). In most of the literature present, a POI represents a static entity, such as hospitals, restaurants, and schools. We use the same idea in this paper. For simplicity, each POI is modeled by its orthogonal projection into the closest edge in $E$ and placed as if it was there. Otherwise, there would be a need to create more nodes and edges to support the new objects.

To quantify the relations between crimes and POIs, we investigate their correlations on the real-world dataset that comprises the crimes in Fortaleza, Brazil, and only those in the theft category. The data contains the following features: the spatial position (latitude, longitude) where the crime was reported and its time. In particular, we first generated a vector to reflect the spatial and time of where and when the crime occurred. We also append the density of the POIs data. To calculate the correlation, we divided Fortaleza City

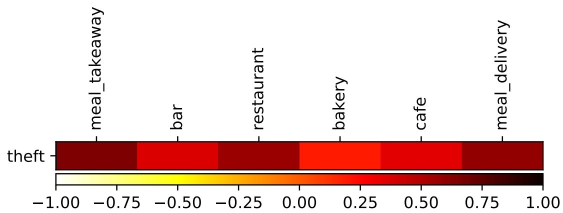



**Figure 2: Correlation analysis between theft occurrences and the density of POIs**

into ten Integrated Security Areas (ISA). Those areas are administrative divisions used in the Public Security State Department that involve some city neighborhoods. For all ten ISA, we counted the number of POIs associated with desired categories and the number of crimes in the area around those POIs. We used those aggregated values to calculate the Person correlation coefficients. Police officers from Fortaleza, Ceará, Brazil chose some categories of POIs for believing in their influence in the theft occurrences. They chose bars, restaurants, and some sustenance facilities, the categories of POIs that are more probable to have a high number of potential victims and offenders visiting the area.

Figure 2 shows the correlation analysis of the dataset as evaluated by Pearson correlation coefficients ($\rho$). This coefficient shows if two variables $X$ and $Y$ are related linearly, assuming values varying from $\rho = -1$ to $\rho = +1$. If $\rho = +1$, that means that the variables $X$ and $Y$ are perfect and positively correlated. If $\rho = -1$, the variables $X$ and $Y$ are perfect and negatively correlated. Lastly, if $\rho = 0$, the variables $X$ and $Y$ do not have any correlation.

From Figure 2, we can observe that all categories of POIs chosen are positively correlated with crime occurrences for the theft category. For instance, considering that $\rho < 0.7$ means a moderated correlation between the two features analyzed, theft is more likely to happen in regions with denser POIs of the meal takeaway category ($\rho = 0.605$) and meal delivery category ($\rho = 0.593$). However, considering that $0.0 < \rho < 0.3$ means a negligible correlation, POIs from the bakery category do not impact thefts ($\rho = 0.188$). We further utilize the POIs data in our predictive policing approach and give the following definitions, which serve as our model's inputs.

We also apply Kernel Density Estimation (KDE) [11], a well-known technique used to create criminal hotspots, to estimate the probability of crime occurs at a target POI based on historical crime occurrences, as shown in Figure 3. The KDE algorithm version works as follows: for each POI, a circle with a 100m radius and center on it is considered, and each POI is scored by the density of crimes inside the described area. We choose the 100m radius because it is the average size of a city block in the region analyzed. The kernel function is built according to each POI score and indicates the estimation of the density of crimes close to the POI. Figure 3 shows one example of POIs in Fortaleza generated with the scores calculated by the *kernel* function, classified in a *choropleth map*. The POI color intensity represents the density of crimes that occurred and were map-matched to it. Notice that the distribution of crime is not uniform. Some areas are dense, while others are sparse. This is already expected since it depends on the opportunity of crime

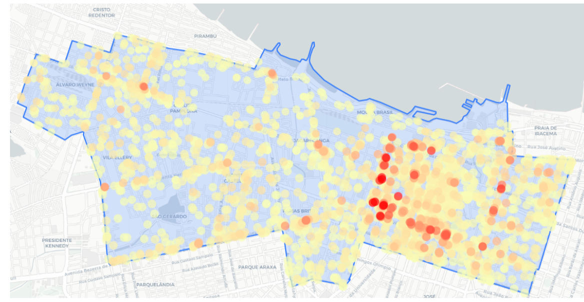


**Figure 3: POIs with different density of crimes.**

and environmental conditions (i.e., location, if it is a busy environment, among others). Figure 3 specifically depicts the downtown of Fortaleza City on the right side. The downtown is an area with lots of POIs, and, as shown in the figure, a high amount of crime is concentrated.

## 2.2 Problem Definition

Let $R = \{r_1, r_2, ..., r_N\}$ denote a set of regions in a city, where $N$ is the number of regions (in our experiments, a region is a circle around a POI). Suppose that there are totally $K$ time slots (i.e., minutes, hours, days), i.e., $T = \{t_1, t_2, ..., t_K\} \in \mathbb{R}^{\mathrm{K}}$. Let $Y \in R^{N \times K}$ be the numbers of crime incidents reported where $Y_n^k$ is the crime number at region $r_n$ in time slot $t_k$.

Now, we can define a function that returns a feature value explored in this paper.

*Definition 2.1 (**Popular time function**).* Let $T = \{t_1, t_2, ..., t_K\} \in \mathbb{R}^{\mathrm{K}}$ be the time slots and $R = r_1, r_2, ..., r_N$ denote a set of regions in a city, the popular time function $popular(r_n, t_k)$ is a number that represents the average popularity over the last several weeks for the region $r_n \in R$ at the time slot $t_k \in T$.

Let $X^k = [X_1^k, X_2^k, ..., X_N^k] \in \mathbb{R}^{\mathrm{N \times K}}$ denote the feature matrix of all regions in the time slot $t_k$, where $N$ is the number of features. One of the features is the popular time (obtained from the function defined above). As we mentioned before, we explored in this paper spatial, temporal, and semantic features. We will discuss all of them in the next section. We can now state our problem we intend to address.

**Problem Statement.** Given the feature matrices $X^1, X^2, ..., X^K$ and the historical crime incidents reported $Y$ of regions in $R$, our goal is to learn a predictive model which estimates the unknown crime occurrence for each region in $R$ in time $t_{K+h}$ ($h$ is a number of time slots) by leveraging $X_1, X_2, ..., X_K$ and $Y$.

# 3 RELATED WORK

This paper exploits spatial, temporal, and semantic features of POIs to build a predictive policing model. Our study is different from [24] since the crime incidents analyzed in our work follows a kind of opportunist behavior. Furthermore, from the information provided in POI data, we exploit their location and temporal and semantic features from these POIs to build a predictive policing model. This is different from previous approaches that use different datasets, e.g., meteorological, human mobility, and 311 public-service complaint data, and do not evaluate them individually.

We divide this section into two groups. The first one discusses the works that propose (semi-supervised) clustering-based models for crime pattern detection, and the second one works on supervised techniques as classification. Some of them also study the influence of spatial and temporal features from crime report data underlying the crime occurrences.

The paper [16] formulates crime pattern detection as a machine learning task. It tries to identify the crime patterns from many crimes, making the job for crime detectives easier by applying a clustering algorithm. The paper [21] proposes Series Finder, a pattern detection algorithm that grows a pattern of discovered crimes from within a database, starting from a seed of a few crimes. The papers [1] and [18] perform a clustering algorithm to discover crime patterns, but in a semi-supervised way, in the sense of using labeled data to generate seed clusters that initialize a clustering algorithm, as well as the use of constraints generated from the labeled data to guide the clustering process.

The second group of techniques is supervised methods. [12] proposes the DeepCrime framework, which enables predicting crime occurrences of different categories in each region of a city by jointly embedding all spatial, temporal, and categorical signals into hidden representation vectors and capturing crime dynamics with an attentive hierarchical recurrent network. [24] exploits temporal-spatial correlations for crime prediction with urban data and proposes a TCP framework that captures temporal-spatial correlations for crime prediction. [9] performs Twitter-specific linguistic analysis and statistical topic modeling to automatically identify discussion topics across a major city in the United States. The authors incorporate these topics into a crime prediction model and show that, for most crime types studied, the addition of Twitter data improves crime prediction performance. [19] observed significantly improved performance in crime rate inference compared to using standard features when it used POI data and taxi flow data. The authors from the paper [22] demonstrate that by dividing the analyzed area into heterogeneous partitions taking into account the density of crime improves crime prediction. The paper explores the heterogeneous division's effect on three prediction methods (Moving Average, ARIMA, and LSTM). [22] models the crime prediction as sequence prediction that involves using sequence information of $q$ values (for instance) to predict the next value.

## 4 DATASET PREPARATION

We used different data sources with relevant geospatial, semantic, and temporal information for the public security domain to build the training dataset. Next, we describe these data sources in detail.

- **Points of Interest**: We collected information on 16,223 Points of Interest located in Fortaleza City from a web mapping service. We extracted the following categories: restaurant, bar, meal takeaway, bakery, cafe, and meal delivery. As we mentioned before, these categories were chosen by the police officers from Fortaleza City because these categories are more likely to have many potential victims and offenders visiting the POI. This fact is in line with the Crime Pattern Theory [5]. Each POI contains information on its name and geographical location (latitude and longitude). As an example, Figure 3 also shows the location of all POIs from the mentioned categories for a delimited region of Fortaleza City.
- **Popular Times**: Some of the collected Points of Interest also have an array of values, with 168 positions, representing its busy hours for every day of a week and every hour in a day (7 days × 24h = 168h). They are called Popular Times. For example, a restaurant has a higher Popular Time value at noon, representing a probable higher number of people having lunch. We also obtained the Popular Times values from a third party API that uses the web mapping service mentioned before. The Popular Times values are calculated based on aggregated and anonymized data from users that frequent the POIs over the last few months. In this work, we are only interested in the POIs that offer Popular Time information.
- **Violent Crimes Against Patrimony**: The Secretariat of Public Security and Social Defense (SSPDS) of the State of Ceará, Brazil, collects the location and timestamp for all Violent Crimes Against Patrimony (CVPs). In general, CVPs are all crimes that did not end up in murder. The SSPDS provided us 115,163 records of CVPs that occurred between 2014 and 2019 in the city of Fortaleza, Brazil. Every record was associated with one crime and has (i) a category of crime (e.g., theft, vehicle, and bank robbery), (ii) geospatial information (latitude and longitude) of where the crime occurred, and (iii) date, and time of the occurrence.

Unfortunately, the Points of Interest and Popular Times dataset can not be publicly available due to the Terms of Services from the used web mapping service. Also, we could not share the Violent Crimes Against Patrimony dataset due to classified information.

Next, we address all the necessary phases to build the training dataset used in the Machine Learning algorithms considered in this work.

**Data cleaning.** In this phase, we chose to use only a fraction of the POIs dataset. We filtered and maintained only the POIs from the bar and restaurant category, resulting in a dataset with 12,178 records. We made this choice because the bar and restaurant category were the ones with the most Popular Times values available. After this first filtering, we joined the new Points of Interest dataset with the Popular Times dataset. We then applied a second filter and excluded all Points of Interest that did not have Popular Times, resulting in a dataset with 2,140 records. We also filtered the CVPs dataset and considered only the theft category since that category is the one with most occurrences. After this filter, the final CVPs dataset contained 62,145 records.

**Data aggregation.** This phase aims to construct a data frame with all necessary pieces of information about the Points of Interest to create a final dataset used to train our models. With that said, we divided our dataset into 168 time slots, representing every hour of the day from one week, starting on Sunday, at 0 o'clock, and finishing on Saturday, at 23 o'clock. Then, for each Point of Interest in the previous dataset and each of the periods, the attributes in Table 1 were obtained or calculated using the CVP's datasets and the Popular Times dataset.

**Table 1: List of attributes aggregated from different datasets.**

| Atribute | Description |
|---|---|
| *name* | The name of the Point of Interest. |
| *latitude* | Latitude of the Point of Interest in the road network. |
| *longitude* | Longitude of the Point of Interest in the road network. |
| *amount_of_crimes* | Number of crimes that happened for each time slot, in a 100m radius from the Point of Interest. |
| *popular_times* | Popular times for each time period for the Point of Interest. |

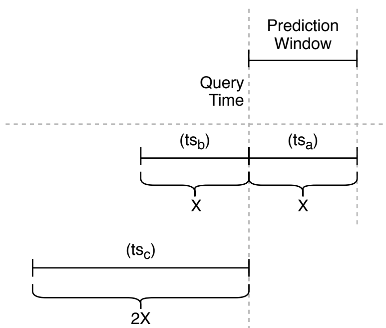



**Figure 4: Diagram showing the engineering of time stamps**

**Data engineering.** After creating all initial attributes at the data aggregation phase, joining pieces of information from the crime occurrences and popular times of the Points of Interest, in the data engineering phase, we also add the temporal dimension to our model. Figure 4 shows how we divided the time domain into intervals. The *Query Time* represents the instant where the prediction request takes place.

The *Prediction Window*, also called $ts_a$, is the interval after the Query Time that we wish to predict how many crimes happen for every POI in a 100 meters radius.

We also use information from the periods before the query time, as follows:

- $ts_b$: time slot with the same size as $ts_a$ that ends also at the same time that $ts_a$ starts.
- $ts_c$: time slot with double the size of $ts_a$ that ends also at the same time that $ts_a$ starts.

To exemplify the previous definitions, consider the following example: for a Query Time equal to 93 (representing Wednesday, at 21 o'clock), we aim to know the number of crimes in a future time slot ($ts_a$) of two hours, i.e., Wednesday, from 21 o'clock to 23 o'clock. For this prediction, we would consider features from the past two hours ($ts_b$), i.e., Wednesday, from 19 o'clock to 21 o'clock, and four hours ($ts_c$), i.e., Wednesday, from 17 o'clock to 21 o'clock. The police resources can be optimally allocated to stop crimes and distributed in different POIs based on the predictions.

With all phases put together, we end up with training sets with 28 features that model the number of crimes in a delimited region around Points of Interest and the time dimension. Table 2 summarizes the features. We decide to profit from the output of some aggregate functions (average, sum, max, and min) over the features. This is commonly used in feature extraction and feature construction, which involve finding a set of *composite* features, which are functions of the original features. According to [23], feature extraction projects a high-dimension feature space to a low-dimension space via linear/non-linear transformations such that most of the information in the original features are retained. Feature construction addresses the problem of feature interaction by discovering good combinations of the original features. We refer the reader to [10, 15] for further details.

We use the $sum\_amount\_of\_crimes\_next$ as the prediction label for our dataset.

The scenarios indicated in Table 2 represent the different combination of features that we aim to model:

- **Scenario 1:** A model created based only on the latitude, longitude, and timestamp features.
- **Scenario 2:** A model created based on the latitude, longitude, timestamp, and popular times features.
- **Scenario 3:** A model created based on the latitude, longitude, timestamp, popular times, and the number of crimes features.

With those scenarios, we aim at evaluating the influence of the different combinations of features (popular times, amount of crimes, and latitude and longitude) and the number of crimes in the model creation. That is what we will study in the next section.

## 5 EXPERIMENTS

This section discusses the experiments conducted to generate different models using different features on the training dataset. More specifically, Section 5.1 describes the experimental setting of the models' generation, and Section 5.3 discusses the results found.

### 5.1 Experimental settings

**Experimental settings.** We conducted all experiments on a Macbook Pro computer with 2,2 GHz Quad-Core Intel Core i7, 16 GB 1600 MHz DDR3, and macOS Catalina (10.15).

**Dataset.** The dataset used in the experiments contains 51,552 records for each of the three scenarios discussed in Section 3 and for each of the four different prediction windows used. The prediction window has the same size as the time interval of $ts_a$ (in other words, the prediction window size is $x$). Therefore, the experiments analyze a total of 12 datasets. We divided the dataset into training (80% of the original dataset) and test set (20% of the original dataset).

**Machine learning methods.** To choose and train the models used in the experimental evaluation, we used the AutoML function from the H2O Machine Learning framework [14]. To create each model, we set the maximum runtime to one hour. H2O uses algorithms like Random Forest, Gradient Boosting Machines (GBMs), and Neural Networks.

### 5.2 Evaluation Metrics

For all the experiments described in this work, we used the Root Mean Square Error (RMSE) to evaluate the models created.

The Root Mean Square Error (RMSE) is given by the Equation 1, where $y_i$ is the predicted amount of crimes and $\bar{y}_i$ is the real value.

**Table 2: Features used to create all models**

| Scenario | Features | Name | Description |
|---|---|---|---|
| 1, 2 and 3 | Point of Interest | latitude | Latitude part of the POI coordinate |
| | | longitude | Longitude part of the POI coordinate |
| | Timestamp | timestamp | Query time that the prediction request occurs |
| 2 and 3 | Popular times X hours before Query Time | sum_populartimes_prevX | Sum of popular times |
| | | std_populartimes_prevX | Standard Deviation |
| | | max_populartimes_prevX | Maximum popular time |
| | | min_populartimes_prevX | Minimum popular time |
| | | avg_populartimes_prevX | Average popular time |
| | Popular times 2X hours before Query Time | sum_populartimes_prev2X | Sum of popular times |
| | | std_populartimes_prev2X | Standard Deviation |
| | | max_populartimes_prev2X | Maximum popular time |
| | | min_populartimes_prev2X | Minimum popular time |
| | | avg_populartimes_prev2X | Average popular time |
| | Popular times X hours after Query Time | sum_populartimes_nextX | Sum of popular times |
| | | std_populartimes_nextX | Standard Deviation |
| | | max_populartimes_nextX | Maximum popular time |
| | | min_populartimes_nextX | Minimum popular time |
| | | avg_populartimes_nextX | Average popular time |
| 3 | Amount of crimes X hours before Query Time | sum_amount_of_crimes_prevX | Sum of crimes |
| | | std_amount_of_crimes_prevX | Standard Deviation |
| | | max_amount_of_crimes_prevX | Maximum number of crimes |
| | | min_amount_of_crimes_prevX | Minimum number of crimes |
| | | avg_amount_of_crimes_prevX | Average number of crimes |
| | Amount of crimes 2X hours before Query Time | sum_amount_of_crimes_prev2X | Sum of crimes |
| | | std_amount_of_crimes_prev2X | Standard Deviation |
| | | max_amount_of_crimes_prev2X | Maximum number of crimes |
| | | min_amount_of_crimes_prev2X | Minimum number of crimes |
| | | avg_amount_of_crimes_prev2X | Average number of crimes |

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n}} \quad (1)$$

## 5.3 Experimental results

For this work, we tried to answer the following questions with the described experiments:

- **Q1**. Which Prediction Window generates the models with the best results, e.g., smaller Root Mean Square Error (RMSE), for the crime occurrence prediction problem?
- **Q2**. Which features are the most relevant for each model created? Also, how the feature relevance change for the different scenarios/features analyzed?
- **Q3**. Which algorithm returns the best model for the crime occurrence prediction problem?

*5.3.1 Q1 – Prediction Window.* This batch of experiments' main goal is to vary the prediction window size $x$ to evaluate which one returns the model with the best results. We used the features described in Table 2 for the POIs category bar and restaurant, separately. For each POI instance of each category, we created a dataset with $168 - 3x$ records, where $x$ is the time window size, in hours. The reason for this number of records for each POI instance is the need to evaluate the features described for the previous $x$ hours and $2x$ hours before the Query Time and $x$ hours after it, as shown in Figure 4. With that said, the dataset for the bar category has approximately 58,000 records, and the dataset for the restaurant category has approximately 288,000 records.

For each dataset created according to the methodology presented in Section 4, we created a training set with 70% of the original dataset and a test set with 30% of the same dataset.

Table 3 show the RMSE measure for each model trained in each of the three scenarios and for each of the six different time windows. According to the results found, smaller time windows provide better results. That is an expected result since by increasing the time window size, more challenging is to build an accurate prediction model. This is in line with the crime theory [7], the occurrence of a crime actively increases the probability of further incidents in the temporal vicinity. The experiments show that the models created are always better for the $x = 2$ time window size and get progressively worst until the $x = 12$ time window. In the scenario of $x = 12$, even with more training data, the models are less accurate to predict the number of crimes because of the large time window size.

For all the results found in Table 3, we see that scenario 3, which involves features related to the number of crimes around the POIs, always delivers the worst results compared to the other scenarios. This fact shows us that those features do not help the algorithm to create more accurate models. This behavior is the same for both categories of POI, i.e., bar and restaurant.

It is worth to mention that XGBoost, which is one of the most used methods by data scientists. In all the considered time window size variations in this batch of experiments, AutoML chose the

**Table 3: Root Mean Square Error (RMSE) for the best models trained for each combination of time window and scenario.**

**(a) bar category.**

| time window | scenario 1 | scenario 2 | scenario 3 |
|---|---|---|---|
| x=2 | **0.7408** | **0.7576** | **1.7229** |
| x=3 | 0.9913 | 0.9889 | 2.0451 |
| x=4 | 1.2225 | 1.1841 | 3.5868 |
| x=6 | 1.6250 | 1.5285 | 3.3682 |
| x=8 | 2.0883 | 2.1126 | 3.7774 |
| x=12 | 2.9690 | 2.9109 | 7.9249 |

**(b) restaurant category.**

| time window | scenario 1 | scenario 2 | scenario 3 |
|---|---|---|---|
| x=2 | **0.5085** | **0.4857** | **0.7998** |
| x=3 | 0.6643 | 0.6356 | 1.0950 |
| x=4 | 0.8120 | 0.7731 | 1.0438 |
| x=6 | 1.1805 | 0.9706 | 1.4891 |
| x=8 | 1.5449 | 1.1593 | 1.8576 |
| x=12 | 2.0570 | 1.5142 | 3.6415 |

model trained using XGBoost as the best one in terms of RMSE. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. We refer the reader to [6] for further details.

*5.3.2 Q2 – Feature relevance.* To answer question Q2, we analyzed the feature importance of the XGBoost algorithm's models in the previous batch of experiments. We used the datasets for the bar and restaurant categories of POIs with a time window of two hours ($x$=2 because we achieved the best results). We also analyzed the three different scenarios described in Section 4 in order to discover the most relevant features. Figure 5 presents the results for the bar category and Figure 6 for the restaurant category.

From the bars plot presented, we can observe that for all scenarios and both categories of Points of Interest, the spatial coordinates from the POIs and the timestamp representing the instant where the prediction request takes place are the essential features for the models created. Those results show us that the location of the Point of Interest and a given timestamp are more critical for predictive policing. Moreover, most of the features related to the popular time information of a POI are more relevant than the historical information about the number of crimes in that region.

*5.3.3 Q3 – Model generation algorithm.* Differently from the batch of experiments used to answer question Q1, in this batch, we chose the time window $x = 2$ that generated the best models so far and ran specific algorithms to generate new models.

The difference from these experiments from the one executed to answer question Q1 is that, previously, we only set the training limit of time to one hour and let H2O AutoML choose the algorithms that should be used. This time, we picked up three different algorithms (XGBoost, GLM, and DRF) and let H2O AutoML train different models, setting a training limit of time to one hour for each algorithm. We did this setup of experiments for the bar and restaurant categories and all three scenarios.

**Table 4: Root Mean Square Error (RMSE) for different algorithms and a time window of two hours.**

**(a) bar category.**

| algorithm | scenario 1 | scenario 2 | scenario 3 |
|---|---|---|---|
| XGBoost | **0.7424** | **0.7587** | 1.7229 |
| GLM | 17.8205 | 22.7485 | 19.5937 |
| DRF | 0.7550 | 0.7667 | **1.5294** |

**(b) restaurant category.**

| algorithm | scenario 1 | scenario 2 | scenario 3 |
|---|---|---|---|
| XGBoost | 0.5085 | **0.4857** | **0.8089** |
| GLM | 7.5647 | 11.8125 | 9.9537 |
| DRF | **0.4736** | 0.5203 | 1.1085 |

According to the H2O documentation, XGBoost is a supervised learning algorithm that implements a process called boosting to produce more accurate models. Boosting refers to the ensemble learning method of building many models sequentially, with each new model attempting to adjust for the deficiencies in the preceding model.

The Generalized Linear Models (GLM) used by H2O AutoML estimate regression models for outcomes following exponential distributions, e.g., normal, Poisson, binomial, and gamma distributions. Each serves a distinct purpose, and depending on distribution and link function choice can be used either for prediction or classification.

Distributed Random Forest (DRF) is a robust tool that generates classification and regression models. When given a dataset, DRF generates a forest of classification or regression trees. Each of these trees is a weak learner built on a subset of rows and columns. More trees will reduce the variance. Both classification and regression take the average prediction over all of their trees to make a final prediction, predicting a class or numeric value.

Table 4 shows the result of this new batch of experiments for the bar and restaurant categories. For all three scenarios, XGBoost outperforms all the other algorithms producing models with smaller errors. It is worth mention that scenario 1, composed only by the latitude, longitude, and timestamp features, presents the best results for the predictive policing problem.

## 6 CONCLUSION

The predictive policing modeling problem should use any critical feature that helps to produce better models. In this context, we consider the use of Point of Interest datasets to better predict the number of crimes in a specific region. We used the geospatial coordinates of a POI, as well as their popular times, as features. We also aggregated data from the number of crimes around the Points of Interest considered. The temporal dimension of the criminal records used was engineered to create new features that aggregated historical information for a previous time window, based on given query time.

To analyze the importance of the popular times of the Points of Interest and the number of crimes around them, different scenarios were tested in the experimental evaluation. We found that the

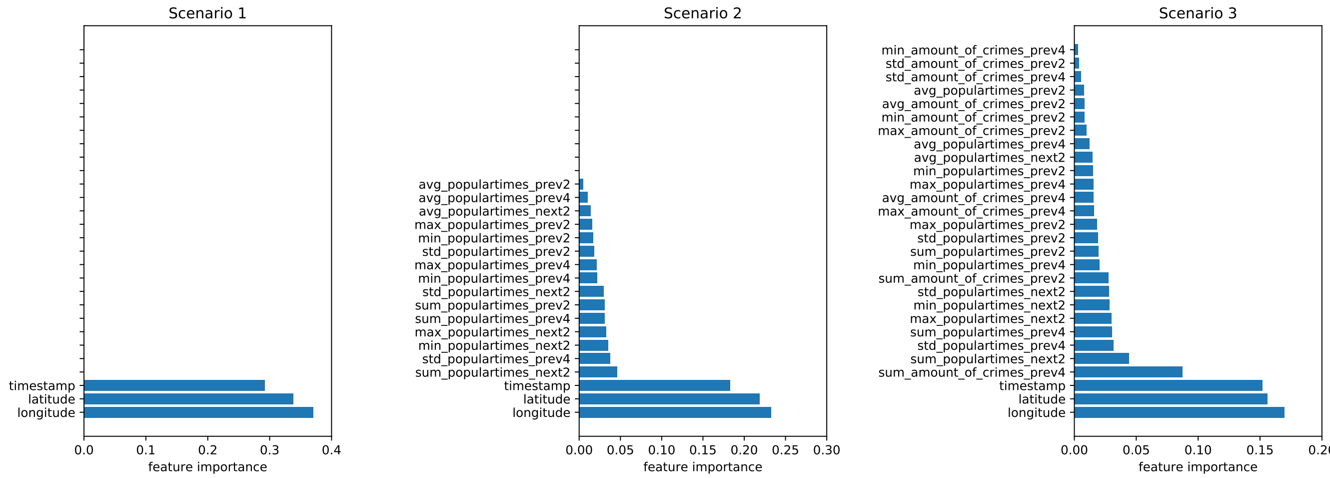



**Figure 5: Feature importance for the bar category of POIs, time window of two hours, and three different scenarios.**

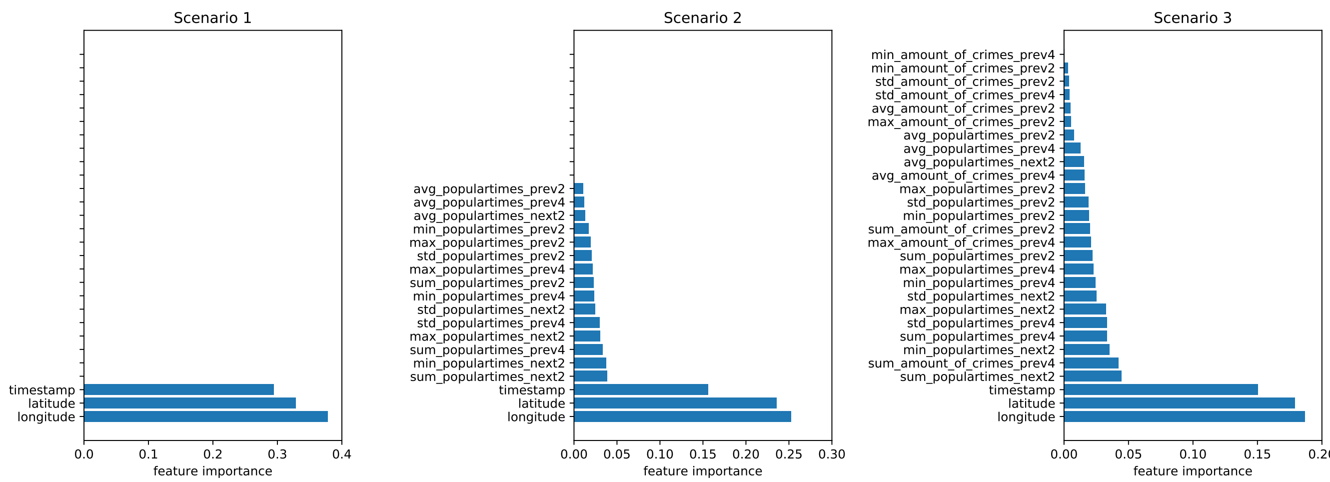



**Figure 6: Feature importance for the restaurant category of POIs, time window of two hours, and three different scenarios.**

addition of the historical amount of crimes does not improve the model quality. Also, all models' essential features are the geospatial coordinates of the POIs and the time when the prediction takes place. When comparing different algorithms, the XGBoosts outperforms all the others, i.e., GLM and DRF.

As future work, we aim to analyze sliding windows in opposition to this paper's fixed ones. Also, we aim at adding to the training set information about the neighbors' Points of Interest, e.g., the number of crimes and their popular times. For the Points of Interest that do not have popular times associated with it, future work could create an estimator to describe those popular times based on Points of Interest with the same category or neighborhood. We will also investigate the POIs, which are risky facilities [4], that is, any group of similar facilities (for which) a small proportion of the group accounts for the majority of crime.

Another line of future work is to develop algorithms to, based on the prediction models created, allocate police resources to prevent crimes. The intuition behind the automatic allocation of the police patrols based on the prediction model's output and the popular time of each POI is the allocation should consider POIs that are spatially close and with complementary popular time.

## 7 ACKNOWLEDGMENTS

This work is partially supported by the FUNCAP SPU 04772551/2020, and the UFC-FASTEF 31/2019.

## REFERENCES

[1] Sugato Basu, Arindam Banerjee, and Raymond Mooney. 2002. Semi-supervised clustering by seeding. In *Proceedings of the Nineteenth International Conference on Machine Learning*. Citeseer, Morgan Kaufmann Publishers Inc., 340 Pine Street, Sixth Floor, San Francisco, CA, United States.

[2] Wim Bernasco and Richard Block. 2011. Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency* 48, 1 (2011), 33–57.

[3] Michail Bourmpos, Apostolos Argyris, and Dimitris Syvridis. 2014. Smart city surveillance through low-cost fiber sensors in metropolitan optical networks. *Fiber and Integrated Optics* 33, 3 (2014), 205–223.

[4] Kate Bowers. 2014. Risky facilities: Crime radiators or crime absorbers? A comparison of internal and external levels of theft. *Journal of Quantitative Criminology*

30, 3 (2014), 389–414.

[5] Patricia Brantingham and Paul Brantingham. 2008. Crime pattern theory. In *Environmental criminology and crime analysis*, Richard Wortley and Lorraine Mazerolle (Eds.). Willan Publishing, 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN, 78–94.

[6] Tianqi Chen and Carlos Ernesto Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Alex Smola, Charu Aggarwal, Dou Shen, and Rajeev Rastogi (Eds.). Association for Computing Machinery, New York, NY, United States, 785–794.

[7] T Cheng, K Bowers, P Longley, J Shawe-Taylor, and T Adams. 2016. *CPC: Crime, Policing and Citizenship – Intelligent policing and big data*. Bantam, UCL Space-Time Lab.

[8] Secretaria de Segurança Pública e Defesa Social. 2019. Criminal Indicators 2018. https://www.sspds.ce.gov.br/estatisticas-2018/. Accessed: 2019-08-29.

[9] Matthew Gerber. 2014. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems* 61 (2014), 115–125.

[10] Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti Zadeh. 2008. *Feature Extraction: Foundations and Applications*. Springer-Verlag, Berlin, Heidelberg.

[11] Timothy Hart and Paul Zandbergen. 2014. Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting. *Policing: An International Journal of Police Strategies & Management* 37, 2 (2014), 305–323.

[12] Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh Chawla. 2018. DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Alfredo Cuzzocrea (Ed.). Association for Computing Machinery, Torino, Italy, 1423–1432.

[13] Etienne G Krug, James A Mercy, Linda L Dahlberg, and Anthony B Zwi. 2002. The world report on violence and health. *The lancet* 360, 9339 (2002), 1083–1088.

[14] Erin LeDell and S Poirier. 2020. H2o automl: Scalable automatic machine learning. In *Proceedings of the AutoML Workshop at ICML*, Vol. 2020. Virtual, -.

[15] Huan Liu and Hiroshi Motoda. 1998. *Feature extraction, construction and selection: A data mining perspective*. Vol. 453. Springer Science & Business Media, USA.

[16] Shyam Varan Nath. 2006. Crime pattern detection using data mining. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops*. IEEE, IEEE, Hong Kong, China, 41–44.

[17] Jerry Ratcliffe. 2005. The effectiveness of police intelligence management: A New Zealand case study. *Police Practice and Research* 6, 5 (2005), 435–451.

[18] Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, et al. 2001. Constrained k-means clustering with background knowledge. In *Icml*, Vol. 1. Morgan Kaufmann Publishers Inc., 340 Pine Street, Sixth Floor, San Francisco, CA, United States, 577–584.

[19] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, Association for Computing Machinery, New York, NY, United States, 635–644.

[20] Hongjian Wang, Huaxiu Yao, Daniel Kifer, Corina Graif, and Zhenhui Li. 2017. Non-stationary model for crime rate inference using modern urban data. *IEEE transactions on big data* 5, 2 (2017), 180–194.

[21] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri. 2013. Learning to Detect Patterns of Crime. In *Machine Learning and Knowledge Discovery in Databases*. Vol. 8190. Springer-Verlag, Berlin, Heidelberg.

[22] Chuanxiu Xiong, Ajitesh Srivastava, Rajgopal Kannan, Omkar Damle, Viktor Prasanna, and Erroll Southers. 2019. On Predicting Crime with Heterogeneous Spatial Patterns: Methods and Evaluation. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Farnoush Banaei-Kashani, Goce Trajcevski, Ralf Güting, Lars Kulik, and Shawn Newsam (Eds.). Association for Computing Machinery, Chicago, IL, USA, 43–51.

[23] Huimin Zhao, Atish Sinha, and Wei Ge. 2009. Effects of feature construction on classification performance: An empirical study in bank failure prediction. *Expert Systems with Applications* 36 (2009), 2633–2644.

[24] Xiangyu Zhao and Jiliang Tang. 2017. Modeling Temporal-Spatial Correlations for Crime Prediction. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Ee-Peng Lim and Marianne Winslett (Eds.). Association for Computing Machinery, Singapore, 497—-506.