Crime Analysis and Visualization in Atlanta Based on Machine Learning Method

Guangyu Min
gmin8@gatech.edu

Ziheng Xiao
zxiao76@gatech.edu

Jingjing Ye
jingjing@gatech.edu

ABSTRACT

As the impact of the crime has increased, predictive hotspot mapping has been widely adopted to effectively curb the occurrence of crime. **In this research**, we will apply a spatio-temporal kernel density estimation (STKDE) method which considers both distance and temporal patterns in crime occurrence distribution to predict crime hotspot. Then the Prediction Accuracy Index (PAI) will be utilized to evaluate the reliability of the predictive hotspots within several area scales.

KEYWORDS

Spatio-temporal kernel density estimation (STKDE), Prediction Accuracy Index (PAI), Optimal bandwidth, Crime hotspot mapping

1 INTRODUCTION

The crime rate in Atlanta has risen sharply, especially for property-related crimes. The prediction or visualization may be needed to help policymakers and relevant departments understand the challenges and problems in the field of crime control system. **Various techniques** of predicting the target hotspots have been used, such as thematic mapping and kernel density mapping. **However**, the existing methods ignore the temporal component which deprives the opportunity of targeting the specific periods with time-related crime incidents.

2 INNOVATION

Our approach provides **time adaptive mechanism** that can import new crime data and update the model. For a specific location, the model will predict the crime rate and type around it. Then based on the prediction, **a suggested destination** that minimizes the risk, will be calculated. **Taxi drivers** can choose their preferred pickup locations according to these crime predictions nearby. **Pedestrians** can plan their route to avoid high-risk regions. With the support of previous models and large data-sets, **a minimum prediction accuracy can be guaranteed**. In addition, we may import noisy testing data to validate the robustness of our model. The possible **impact** is to provide a possible route for drivers and pedestrians. It may also **contribute to the crime analysis of organizations** if the model reaches the accuracy threshold.

One **difficulty** of our model is to define the risk level of a region with accuracy, based on the prediction. A safe location often can be predicted as high-risk, with heavy weights around it. Another **difficulty** is the time adaptive mechanism that fits the model when new crime data is imported. To make the model reliable, the updating process must be sensitive enough to the data. The **payoff** is to make real-time data acceptable and increase the prediction accuracy.

3 LITERATURE SURVEY

To do the crime analysis, we do surveys on several ways. One direction is Hotspot mapping.[3]

Hotspot mapping has become a popular analytical technique used by law enforcement, police, and crime reduction agencies to visually identify where crime tends to be highest. This paper analyzes the Spatial ellipses, Thematic mapping of geographic boundary areas, Grid thematic mapping, and Kernel density estimation. Finally, get the conclusion that KDE performs best, but KDE requires two parameters - the cell size and the bandwidth which has an influence on hotspot maps.

Another direction is using neural networks like ANN. Corcoran et al.[5] introduced a crime prediction model for high-risk areas using ANN. The model combines with a crime incidence-scanning algorithm [11] to identify clusters of high-risk regions, i.e. hotspots.

Due to the versatility of KDE, more survey has been put here. Standard KDE only include space feature, but temperature, time and many factors affect the prediction. [13] The Prospective Mapping is a direction that can solve the problem. Another thought is to add time dimension to KDE named spatio-temporal kernel density estimation which has been mention by Hu and Nakaya [9][12].

To process the data, Chen et al. [4] use a data mining method identifying subgroups and key members in criminal networks. This illustrated a way to identify interaction patterns between crimes within a crime network. Beshah [1] provided the performance of traditional classifiers in road traffic accident data, which can be used to links recorded crime characteristics to regions.

As bandwidth is important in STKDE model and our framework is a data-driven optimization process, a likelihood cross-validation method can help detect the most appropriate bandwidths, reflecting the distribution trends in the data [14].

On the term of evaluation, Kattan et al.[10] and Chainey et al.[3] choose Prediction Accuracy Index to measure the correctness and reliability of risk prediction models including crime probability and survival probability.

To do better, the model can be time adaptive. Ricardo et al.[2][7] declares a time adaptive method which imports the Nadaraya-Watson estimator and forgetting factor λ to do conditional kernel estimation on wind power forecasting.

4 TECHNICAL METHODOLOGY

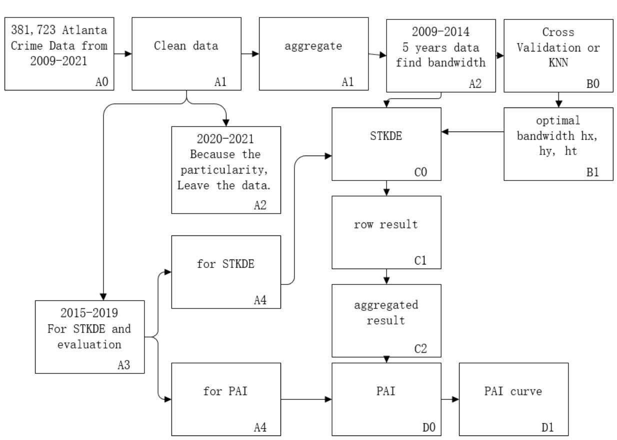

Fig. 1 Methodology diagram

4.1 Spatio-Temporal Kernel Density Estimation

Spatio-Temporal Kernel Density Estimation is a temporal extension of the traditional kernel density estimation KDE used for identifying spatiotemporal patterns. The density estimates are visualized within the space-time cube framework using two-dimension spatial (x, y) and a temporal dimension (t). After running STKDE, it will provide a 3D raster volume as output where each voxel is assigned a density estimate based on the surrounding point data. The space-time density is estimated by the following equation.[8]

$$\widehat{f}(x,y,t) = \frac{1}{nh_s^2 h_t} \sum_i i\,(d_i < h_s, t_i < h_t)\, k_s \times \left(\frac{x-x_i}{h_s}, \frac{y-y_i}{h_s}\right) k_t \left(\frac{t-t_i}{h_t}\right) \quad (1)$$

Density of each voxel s with coordinates (x, y, t) is estimated based on neighboring data points (x_i, y_i, t_i). Each point located within the neighborhood of the voxel is weighted using the spatial and temporal Epanechnikov [6] kernel functions, k_s and k_t, respectively (the closer the data point, the higher the weight). The spatial and temporal distances between voxel and data point are given by d_i and t_i respectively. If d_i and t_i are smaller than the spatial h_s and temporal bandwidths h_t respectively, the indicator function $I(d_i < h_s; t_i < h_t)$equals 1, otherwise 0. The following diagram provides the algorithm.

```
ALGORITHM STKDE(xyzList, hs, ht)  #Inputs: array of coordinates, s/t bandwidths
BEGIN ALGORITHM
for (xC=xmin;xC<=xmax;xC+xRes):                  #for all x-coordinates
   for (yC=ymin;yC<=ymax;yC+yRes):               #for all y-coordinates
      for (zC=xmin;zC<=zmax;zC+zRes):            #for all  t-coordinates
         density = 0.0
         for xD, yD, zD in xyzList:              #loop through disease cases
            if hs >= (xD - xC)^2 + (yD - yC)^2:  #if within spatial bandwidth
               if ht >= |zD – zC|:               #if within temporal bandwidth
                  u = (xD-xC) / hs
                  v = (yD-yC) / hs
                  w = (zD-zC) / ht
                  density += 0.5 * π * (1 – u^2 – v^2) * 0.75 * (1 – w^2)
END ALGORITHM
```

Fig. 2 Sequential STKDE algorithm

4.2 Bandwidth Selection

There are some equations to determine the value of bandwidth, such as plug-in methods, data-driven methods and bootstrap methods. Here we picked the data-driven methods called Least Squares Cross-Validation

(LSCV) [8], which is calculated in the following equation.

$$CV(h,\lambda) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_{h,ij}^{(2)} - 2\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{h,ij}^{(2)} \quad (2)$$

For each picked bandwidth λ, we calculated the kernel density for each voxel and used the equation to determine the quality of each bandwidth value. The criteria is similar to the MSE (Mean Square Error) and ISE (Integral Square Error).

4.3 Prediction Model

To serve the purpose of prediction, we utilize the model which is denoted as Multilayer perceptron (MLP).

A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN).It consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can also distinguish data that is not linearly separable.

4.4 User Interface Design

On the website page, we used Google Maps JavaScript API to import a real map as a background. The city are partitioned into neighborhood areas and denoted with grey color. We used the heatmap to show the crime analysis and prediction for the whole city, as a visualization solution. There are several interaction options:

1. On the bottom slider, draw the bar to select a date you want to view analysis.
2. On the dropdown button, select the hour to view details.
3. Click on the city map (grey region) to see the crime analysis and prediction for a specific date and hour period that are chosen. You may either click inside or outside the city map. When you click outside the map, there probably will be N/A for the prediction result. When you choose a previous datetime, the aggregated information are given in an info window. When you choose a future datetime, the prediction result will be shown.

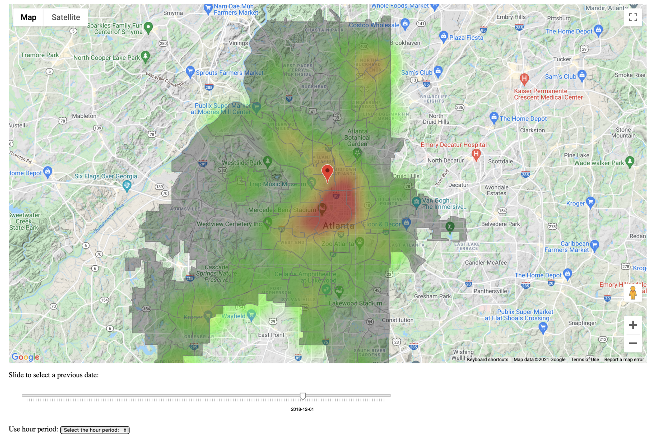

Fig. 3 Crime Map User Interface

5 PLAN OF ACTIVITIES

The above table shows our timelines. Revised part is in bold. For the first stages, crime prediction will be performed based on the 11-years crime data provided by Atlanta Police Department and test it with the data of this year (2021), which takes 3-4 weeks. We've focused on **implementing STKDE method to estimate the probability of crime in both spatial and temporal dimensions simultaneously**. All team members have contributed similar amount of effort, around **33 % contribution** per member.

For the next two stages, it will cost another 2 weeks to first **evaluate the method** and then do the visualization on the interactive web page to show crime hotspots at different time on the map.

For the final stage, it will take 2 weeks to **finalize the evaluation and visualization**. We will **finish all the deliverables** and every member is supposed to have a similar amount of contribution during this stage.

6 DATA

Raw crime data is retrieved from Atlanta Police Department website (Crime Data) and Geographical mapping data will be provided by Atlanta Department of City Planning (Geographical Mapping Data).

The crime data consists of several hundreds of thousands of records ranging from 2009 to 2021 with 19 variables, covering key information of a certain crime like report id, type of cases, occurrence time, precise occurrence location (neighborhood, NPU, longitude and latitude), etc.

Geographical mapping data will be used for interactive visualization of crime prediction as well. Such data are fetched from GIS Resource provided by Atlanta Department of City Planning (Atlanta Geographical Mapping Data) in the form of GeoJSON files.

7 EXPERIMENT

7.1 Data Cleaning

(1). Merged data from csv files retrieved from open data source provided by Atlanta Police Department, which are named as "COBRA-2009-2019.csv", "COBRA-2020-OldRMS-09292020.csv", "COBRA-2020(NEW RMS 9-30 12-31).csv", "COBRA-2021.csv", representing data collected in different years respectively.
(2). Selected several columns which should be significant to the results of prediction and visualization, such as "offense_id", "occur_time", "location", "UCR_Literal", "lat"(latitude), "long"(longitude), etc.
(3). Modified columns of time of occurrence and converted the combined strings to the standard time string as well as time stamps.
(4). Selected a bottom-left point to the main domain of Atlanta city as the origin, modified columns of longitude and latitude of the reported location and converted to the expression in kilometers.
(5). Selected time of 00:00 am on January 1st 2009 as the origin and re-scaled the time stamps to a reasonable range of value.
After data cleaning, 381225 records of crimes remained in the data file.

7.2 Density Calculation

(1). Read the data as x, y, t, where x and y x,y comes from the latitude and longitude information and t is the time stamp for crime occurred.
(2). Set the spatial bandwidth h_s and temporal bandwidth h_t from experiments. The bandwidth defines how does one crime accident care about its neighbors.
(3). Use bandwidth and resolution to define grids and build KDTree to find the neighbors for grid point (x_c, y_c, z_c).
(4). For each neighbor point, apply the equation (1) to calculate the density result and save the density.

The preliminary result by prediction over timestamps divided by 300 generated by the STKDE algorithm can be shown as the following figure, as the redder region representing higher probability of crime occurrence while the bluer region representing the opposite.

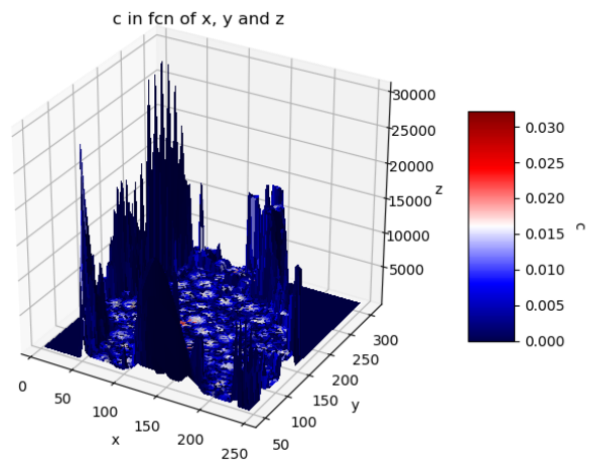

Fig. 4 Probability distribution of crime occurrence over time generated by STKDE algorithm

7.3 Visualization of STKDE

Set h_s and h_t to S m and T days, which was determined preliminary analysis of the space-time kernel density function. We used a spatiotemporal voxel-resolution of S m * S m * T days within our experimental treatments. We plan to present the probablity with scale on a 2-D heatmap. One can select the time range for different seasons or periods in one day.

7.4 Extract Feature on STKDE Output

To get the model with a better performance, five features are selected, which are output data gained by respectively calculating the average probability density of 10 nearest neighbours (distance ranging from 1-10 grids), 20 next nearest neighbours (distance ranging from 11-30 grids), 30 next nearest neighbours (distance ranging from 31-60 grids) , the closest 6 days (including the forward and the backward) and the records of the same day of week.

As is shown in the correlation analysis, features of average probability calculated from the 10 nearest neighbours has a best correlation with the target, which means benefit to the final prediction model, while the feature collected from the closest 6 days shows the worst.

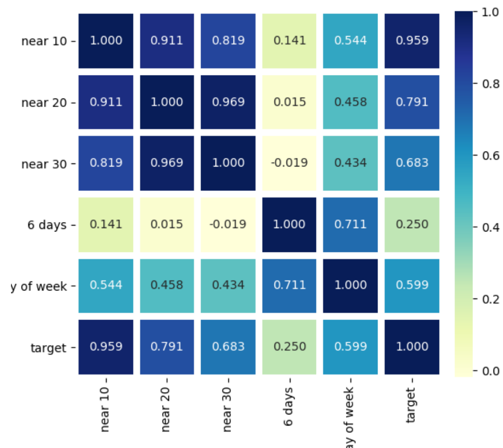

Fig. 5 Correlation heatmap of the selected 5 features

7.5 Predict with Multilayer Perceptron

7.5.1 MLP:. [15] The learning algorithm of MLP is based on the minimization of the error function defined on the learning set (x_i, d_i) for $i = 1, 2, \ldots, N$ using the Euclidean norm:

$$E(w) = \frac{1}{2}\sum_{i=1}^{N}\left\|y\left(x_i, w\right) - d_i\right\|^2 .$$

Adaptation of weights is performed step by step

$$w(k+1) = w(k) + \gamma p(k),$$

where $p(k)$ is the direction of minimization in kth step,γ is the learning coefficient, and w is the adaptation coefficient.

Most effective is the Levenberg–Marquard algorithm for medium size networks and conjugate gradient for large size networks.

Levenberg–Marquard algorithm

Least square formulation of learning problem is exploited:

$$E(w) = \frac{1}{2}\sum_{i=1}^{M}\left(y_i(w) - d_i\right)^2 .$$

Solved by using second order method of Newton type:

$$p(k) = -G(k)^{-1}g(k),$$

where $g(k) = \frac{\partial E}{\partial w(k)}$ is the gradient of error function Eq. $G(k)$ is the Hessian approximation, determined by applying the Jacobian matrix $J(k)$:

$$G(k) = J(k)^T J(k) + v.$$

In this equation the Jacobian matrix J is equal

$$J = \frac{\partial e}{\partial w}e = \left[y_i(w) - d_i, \ldots, y_M(w) - d_M\right]^T .$$

Conjugate gradient

Direction $p(k)$ is evaluated according to the formula.

$$p(k) = -g(k) + \beta p(k-1),$$

where the conjugate coefficient β is usually determined according to the Polak-Ribiere rule:

$$p(k) = \frac{g(k)^T(g(k) - g(k-1))}{g(k-1)^T g(k-1)}.$$

7.5.2 Training process: Put the five features extracted features in to the MLP network. There is the parameter of the training process.

- n = 5
- Split dataset: test size is 0.3.
- MLP: Set three hidden layer(Linear) and use Rectified Linear Unit (ReLU) as the activation function.
- Dataloader: batchsize: 64
- Optimizer: Use Adam as the optimizer and use MSELoss to calculate loss.
- Learning rate: 0.01.
- Epoch: 200.

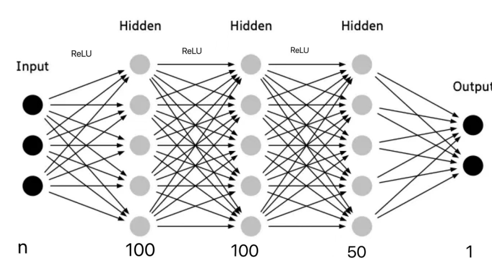

Fig. 6 MLP net

8 EVALUATION

Since STKDE is a kind of unsupervised learning process, after results which indicate the probability of crime in certain areas at a specific time are generated by STKDE, performing an evaluation is needed to test the correctness and the effectiveness of this algorithm.

The first step is to **perform visualization of calculated crime probability** mapped on the geographical layer. In this period, we will be able to check the mapping errors and figure out if there is something counterintuitive (e.g. an area with a fewer number of cases somehow cast a higher probability of crime).

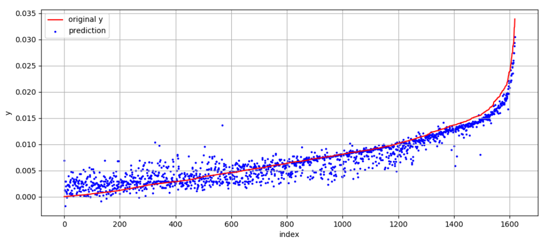

Fig. 7 Distribution of prediction result

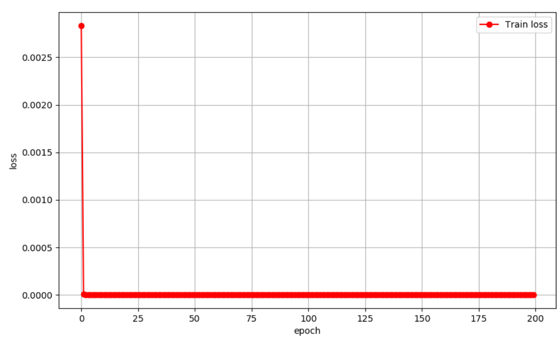

Fig. 8 Decrease of training loss

The second step is to **utilize the Prediction Accuracy Index**, which is frequently applied and useful to evaluate the reliability of risk prediction models, to **measure the accuracy** of our crime prediction model generated by STKDE algorithm.

$$PAI = \frac{HitRate}{AreaPercentage} = \frac{n}{N} \cdot (\frac{a}{A})^{-1},$$

- n, number of events fall in predicted area;
- N, number of total events;
- a, sum of predicted areas; - A, sum of total areas.

As shown in the **Fig.7** and **Fig.8**, during the training process of MLP, the training loss decreased continuously and reached a very small value (2.8531×10^{-6}) finally. The prediction results were very close to the ground truth we set up in the test dataset.

What will cost the most are the **calculation time**, which is also what we concern. We will test the running time cost to perform a one-year prediction based on the previous 11-year crime dataset and optimize the algorithm to the least time consumption. On the other hand, we plan to do the visualization with pre-calculated data in the interactive web page to minimize the time cost.

9 CONCLUSIONS AND DISCUSSION

In this project, we applied a spatio-temporal kernel density estimation method to process 10-year crime data at Atlanta and used model to predict the crime probability for a specific datetime at a location grid. The STKDE method will calculate the density of data in three-dimensional space including space and time. It can generate the analysis for the previous crime data, which serves as the input for the prediction model. The model generated a crime probability value for each grid, which is visible on the crime map demo. The demo will display a heatmap for crime accidents. For each location, it allows users to choose either previous aggregated crime information or the prediction for a future timestamp. With the prediction accuracy index (PAI), the prediction can be further validated and improved.

There are two main possible improvements in our implementation. First, the prediction model can be improved by carefully designing the cross-validation. Currently our data are almost fully used in prediction and the test part are limited. We may take other attributes into consideration as well, such as the crime type, neighborhood information. This may improve the prediction accuracy, but lead to some loss of generality for the model.

Another point is to improve the visualization part. Frankly speaking, our UI functions can be enriched. For example, the heatmap presentation can be turned into figures of clustered markers, which shows the degree of crime accidents in each neighborhood. In addition, We can perform more experiment and generate evaluation results to convince that our model works well.

REFERENCES

[1] Tibebe Beshah and Shawndra Hill. 2010. Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors

on Accident Severity in Ethiopia. *AAAI Spring Symposium - Technical Report*.

[2] Ricardo J. Bessa, Vladimiro Miranda, Audun Botterud, Jianhui Wang, and Emil M. Constantinescu. 2012. Time Adaptive Conditional Kernel Density Estimation for Wind Power Forecasting. *IEEE Transactions on Sustainable Energy* 3, 4 (2012), 660–669. https://doi.org/10.1109/TSTE.2012.2200302

[3] Spencer Chainey, Lisa Tompson, and Sebastian Uhlig. 2008. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal* 21, 1 (01 Feb 2008), 4–28. https://doi.org/10.1057/palgrave.sj.8350066

[4] H. Chen, W. Chung, J.J. Xu, G. Wang, Y. Qin, and M. Chau. 2004. Crime data mining: a general framework and some examples. *Computer* 37, 4 (2004), 50–56. https://doi.org/10.1109/MC.2004.1297301

[5] Jonathan Corcoran, Ian Wilson, and J. Ware. 2003. Predicting the Geo-Temporal Variations of Crime Disorder. *International Journal of Forecasting* 19 (02 2003), 623–634. https://doi.org/10.1016/S0169-2070(03)00095-5

[6] V. A. Epanechnikov. 1969. Non-Parametric Estimation of a Multivariate Probability Density. *Theory of Probability & Its Applications* 14, 1 (1969), 153–158. https://doi.org/10.1137/1114019 arXiv:https://doi.org/10.1137/1114019

[7] Andrew Harvey and Vitaliy Oryshchenko. 2012. Kernel density estimation for time series data. *International Journal of Forecasting* 28, 1 (2012), 3–14. https://doi.org/10.1016/j.ijforecast.2011.02.016 Special Section 1: The Predictability of Financial Markets Special Section 2: Credit Risk Modelling and Forecasting.

[8] Alexander Hohl, Eric Delmelle, Wenwu Tang, and Irene Casas. 2016. Accelerating the discovery of space-time patterns of infectious diseases using parallel computing. *Spatial and Spatio-temporal Epidemiology* 19 (2016), 10–20. https://doi.org/10.1016/j.sste.2016.05.002

[9] Yujie Hu, Fahui Wang, Cecile Guin, and Haojie Zhu. 2018. A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. *Applied Geography* 99 (2018), 89–97. https://doi.org/10.1016/j.apgeog.2018.08.001

[10] M.W Kattan and T.A. Gerds. 2018. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. *Diagnostic and Prognostic Research* 2, 7 (2018). https://doi.org/10.1186/s41512-018-0029-2

[11] Nafiz Mahmud, Khalid Ibn Zinnah, Yeasin Ar Rahman, and Nasim Ahmed. 2016. Crimecast: A crime prediction and strategy direction service. In *2016 19th International Conference on Computer and Information Technology (ICCIT)*. 414–418. https://doi.org/10.1109/ICCITECHN.2016.7860234

[12] Tomoki Nakaya and Keiji Yano. 2010. Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics. *Transactions in GIS* 14, 3 (2010), 223–239.

[13] Gabriel Rosser, Toby Davies, Kate J. Bowers, Shane D. Johnson, and Tao Cheng. 2017. Predictive Crime Mapping: Arbitrary Grids or Street Networks? *Journal of Quantitative Criminology* 33, 3 (01 Sep 2017), 569–594. https://doi.org/10.1007/s10940-016-9321-x

[14] Alexander Suhre, Orhan Arikan, and Ahmed Enis Cetin. 2016. Bandwidth selection for kernel density estimation using Fourier domain constraints. *IET signal processing* 10, 3 (2016), 280–283.

[15] E.A. Zanaty. 2012. Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification. *Egyptian Informatics Journal* 13, 3 (2012), 177–183. https://doi.org/10.1016/j.eij.2012.08.002