International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 3127-3132

Forecasting Violent Crime Hotspots Using a Theory-Driven Algorithm

Falade Adesola, Ambrose Azeta, Aderonke A. Oni, Felix Chidozie

Department of Computer and Information Sciences, Covenant University, Ota, Nigeria.

ORCID: 0000-0001-7142-7405 (Falade)

Abstract

Crime hotspot forecasting is an important part of crime prevention and lessens the delay between a Police call and the physical intervention. Recent researches and developments in this area focus on enriching the historical data and sophisticated point process analysis methods with a fixed grid. In this article, a simple spatio-temporal point process which allows one to perform exhaustive (literal) grid searches is presented. It further show that this approach can compete with other complex methods in literature, as evidenced by the results on data collected by the statistics department of Nigerian Police. Finally, this study discusses the advantages and potential implications of this new method.

Keywords: Predictive Accuracy Index (PAI), Prediction Efficiency Index (PEI), Area Under the Curve (AUC), Receiver Operating Characteristics (ROC) curve. National Institute of Journalism (NIJ).

1 INTRODUCTION

Spatio-temporal crime forecasting is a field that grabs the attention of both scientists and practitioners. Many academic researchers have published results based on time series analysis [7], regression methods [4], [10], [23], kernel density estimation [2], [3], [5], [8], [19], [1] or self-exciting point processes [11], [22], [21], [15], [13], [16], [12]. Moreover, the US Government appreciates the impact predictive policing has on society [18]. In a typical crime prediction task, the forecast area is fixed and divided into small sub-regions, called cells. The cells are then scored separately over a given future time window. The ones with the highest rate are chosen as the most dangerous areas and are called hotspots. In this article, one point of view for hotspot forecasting that differs from those which can be found in literature is presented. However, emphasis is laid on the simplicity and efficiency of the chosen algorithm for a fixed grid to get an opportunity to check as many grids as possible. These attributes are placed over sophisticated methods, with state-of-the-art results in practice.

The rest of the paper is organized as follows. In Section 2, literature review on which the research is premised is discussed. Section 3 presented the approach use for the study and also contains a comprehensive description of case study of our method. Further comments and summary as well as discussion and the results are presented in 4 and 5 respectively.

2 LITERATURE REVIEW

Crime hotspots mapping has been used in literature as the basic form of crime prediction, heavily relying on previous crime dataset to discover the areas of high concentration of criminal activities where Police authority could deploy their limited resources in combatting crime. As previously demonstrated in literature [2], there are three basic crime mapping techniques and these are: Point mapping, Thematic mapping and Kernel Density Estimation (KDE).

According to the research work carried out by [3], Kernel Density Estimation technique was used and applied on crime dataset for a period to generate crime hotspots maps and then test the accuracy of prediction when the next crime will occur. The researchers then compared the accuracy of the crime hotspots mappings in relation to the mapping techniques used in order to identify high concentration of crime events. [3] further showed that KDE is the technique that consistently outperform other crime mapping techniques. However, the results are limited in crime hot spots literal grid searches.

The study carried out by [6] titled “Scalable high-resolution forecasting of sparse spatiotemporal events with Kernel methods: A winning solution to the National Institute of Journalism Real-time crime forecasting challenge” proposed a generic spatiotemporal events forecasting method for NIJ real-time crime forecasting challenge. The researchers applied Kernel Density Estimation and Self-exciting point process models to make spatiotemporal point pattern prediction. The results of the prediction when evaluated with other baselines showed a better performance but lacks capacity in crime hot spots literal grid searches.

[11] proposed a criminal incident prediction using a point pattern based density model for spatiotemporal event prediction. The researchers introduced a multi-variate prediction model for hotspots that relates the features in an area to the predicted crime occurrence through preference structure of criminals. The evolved model outperforms the current practices as evidently and statistically demonstrated. However, the limitation of the model is that it could not able to perform crime exhaustive grid searches.

3 THE MODEL

There is a vast literature available about crime forecasting for a given grid of cells based on past crimes committed. In such a setup, more or less sophisticated methods are applied to predict which fixed parts of the investigated region will

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 3127-3132

experience the highest future rate of crime. Clearly, changing the grid changes the entire task as well and may lead to completely different predictions with different levels of effectiveness in the real world. The choice of grid is really important. However, to the best of our knowledge, whenever the cell division is not imposed in advance, searching for a good grid is in practice reduced to grid search, random search [20] or another primitive method of walking among parameterizations of possible tessellations. The distribution of crimes committed in urban areas are 'weird': they contain atoms with very high crime rates (related to, for example, large-area stores or shelters for the homeless). Therefore, using the same data-driven algorithm for even very similar grids can cause a huge discrepancy in the qualities of the predictions obtained. Hence, grid optimization cannot be neglected. However, a good grid parametrization should take into account horizontal and vertical shifts, cell height, width, rotations, shape distortions, grids based on triangles, hexagons, and aperiodic tessellations. Taking into consideration the massive number of grids worth checking, it was concluded that there was a need for a very fast but still well performing supervised model for a fixed grid, one that would simply execute a random search on a rich space of grid parameterizations to find the 'optimal' grid. This would yield a better final result than a more sophisticated, but slower algorithm applied to a random set of grids that would be too small to contain any decent tessellation.

3.1 Fast Algorithm for a Given Grid

The main idea behind the proposed algorithm for a fixed grid is simple: count the past crimes in every cell and mark the cells with 'the worst past' as hotspots. In other words, it is assumed that if many crimes occurred somewhere, more are likely to happen. This principle may strike some as naive and outdated, but it is believe that it is both accurate enough and fast. Up-to-date crime reports are freely available for several Nigerian cities. They form the main dataset in data-driven crime forecasting algorithms. One can search for any external data which could affect future crimes, but have not left a trace on those crimes that have already been committed. We are aware that weather, demographics and even social media information [23] are sometimes used in similar contexts. Unfortunately, they significantly increase the model's complexity, often without a guarantee of noticeably improved accuracy. Keeping computations as simple as possible, by using merely historical crime data, enables us to spend more time on selecting the right grid.

The raw algorithm is refined by taking care of data aging and seasonality, that is, we assign weights to all the past crimes and then sum up the weights of all the crimes in consecutive cells to find the hotspots. The weight of an event decreases exponentially as a function of age (in days) of a crime. The intensity of the decrease is a hyper-parameter, tuned with the use of available data to obtain the best results. Also, we boost the weights of crimes committed on the same days of the year as those in the forecasted time span. The power of boosting is a hyper-parameter as well. Moreover, we introduce a primitive 'spatial radiation' of past crimes [24], [25]. For each data point, we put eight of its copies with reduced weights in the corners and in the center of the sides of the rhombus around it. In this way, a 'part' of an event that has occurred close to the cell border could fall into a neighboring cell. We chose to use a rhombus because it reflects the Manhattan metric, a reasonable match for North-South-oriented axis grid street plans, of which there are many of such in Nigerian cities. In our opinion, this 'degenerated spatial decay' technique is pretty fast and good enough for working with aggregations of crimes to regular convex cells. The size of the rhombus and reduction of weights of added copies are hyper-parameters [26].

The strict mathematical description of the presented approach, expressed in the language of spatio- temporal point processes [15] is hereby presented.

Given space time-point process N(t, x, y) which is usually symbolize by its given conditional intensity λ(t, x, y), and which can be described as the expected limiting rate of the accumulation of points around a particular spatiotemporal region, given H_t as the history of all points up to time t [12].

$$\lambda(t,x,y) = \lim_{\Delta t,\Delta x,\Delta y \downarrow 0} \Big(E\big[N\{(t,t+\Delta t) \times (x,x+\Delta x) \times (y,y+\Delta y)\} | \mathcal{H}_t\big] / (\Delta t \Delta x \Delta y) \Big). \quad (1)$$

Given data point $(t_k, x_k, y_k)_{k=1}^{N}$ and a self exciting point process model in the form

$$\lambda(t,x,y) = \sum_{t-T<tk<t}^{\infty} g(t-tk, x-xk, y-yk) \quad (2)$$

$$\lambda(t,x,y) = v(t)\mu(x,y) + \sum_{t-T<tk<t}^{\infty} g(t-tk, x-xk, y-yk) \quad (3)$$

Iterations are done until convergence using the following:

Step 1: Given background events $\{(t_i^b, x_i^b, y_i^b)\}_{i=1}^{N}$ and parent/offspring inter-point distances $\{(t_i^0, x_i^0, y_i^0)\}_{i=1}^{N}$ from P_{n-1}.

Step 2: Estimate v_n, μ_n and g_n from the sample data.

Step 3: Update P_n from v_n, μ_n and g_n using (2) and (3).

Therefore, estimating v_n, μ_n and g_n from the sampled data, variable bandwidth Kernel is used in density Estimation. Also to estimate g_n, the data$\{(t_i^0, x_i^0, y_i^0)\}_{i=1}^{N}$ *is scaled* to be able to have unit variance in each coordinate. The rescaled data is used to compute D_i which is the k^{th} nearest neighbor distance (three-dimensional Euclidean distance) with respect to t_i data point. The data is then transformed back to its original scale and by letting σ_x, σ_y, and σ_t be the sample standard deviation of each coordinate.

The simulation was achieved by first simulating all the background events in accordance with Poisson process vμ. The other part of the simulation was iteratively carried out, where each point of each generation produces its own offspring in accordance with Poisson process g centered at the

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 3127-3132

parent point. The entire process came to an end at the n^{th} generation when all the events of the n^{th} generation lie outside the time window under consideration as graphically illustrated Figure 1.

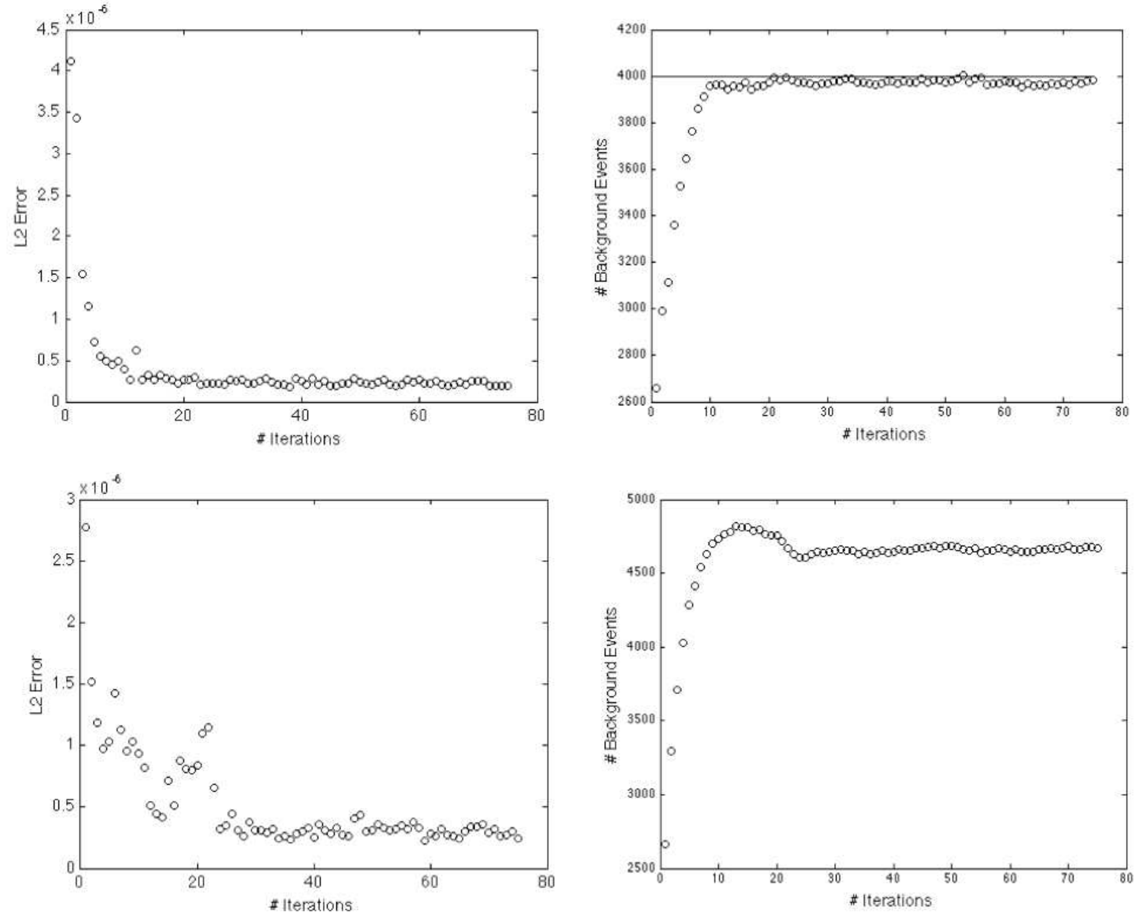

Figure 1: L2 error $\{P_n - P_{n-1}\}$ (top left) and N_b, the number of sampled background events, (top right) at the nth iterations for known point process model. L2 error $\{P_n - P_{n-1}\}$ (bottom left) and N_b, the number of sampled background events, (bottom right) at the nth iterations for the method applied.

3.2 VALIDATION

In classic crime forecasting, the score functions taken from the binary classification –ROC/AUC, sensitivity, etc. – are used [3]. There are also two newer functions from literature: predictive accuracy index (PAI, [3]) and prediction efficiency index (PEI, [9]). They all have their disadvantages. Binary classification-based functions are inconvenient if the area of the hot-spots to be forecast is a very small fraction of the investigated jurisdiction, which is typical. As for other functions, PAI favors smaller single cell areas while PEI likes as great a single cell area as possible.

For this reason it is impossible to maximize both PAI and PEI with the same grid, which casts doubt on the validity of using either of them. Moreover, PEI is bounded by 1 from above whereas the range of PAI is a positive half line, so they are not directly comparable. Here the lack of a simple universal unbiased score function becomes evident. Nevertheless, our approach is metric-agnostic, therefore any reasonable score function can be applied here.

4 DATA

The Statistics department of Nigerian Police delivered historical data on all the crimes registered in Lagos between March 2012 and February 2017. Almost 580,000 records were provided in total. Each of them contained, crime type, description, number of deaths, the day the crime was committed, coordinates (with accuracy to one foot). There were no data gaps. A very small portion of data was located outside the competition area. The distribution of data between crime categories was highly imbalanced: arm robbery, burglaries, car thefts and rape were only 12%, 3%, 7%, and 6.5% of records, respectively. One would anticipate a similar distribution reflected in crimes committed between March and May 2017. Thus, a huge discrepancy is expected in the numbers of crimes committed between particular type/time categories during that period. That was true, two extreme cases were: all the crimes between March and May 2017 - 65,000 records, and burglaries in the first week of March 2017 - only 20 events. Distributions of crimes in all the categories with a big enough number of events had similar characteristics: they consisted of the ’dense’ part looking like a sample from a continuous distribution and the ’discrete’ part made from atoms. It seems that although the accuracy of the coordinates of crimes committed was in general one foot, police officers tended to ’discretize’ some areas like stores or shelters to a single spatial point next to the entrance to the building/area.

4.1 Computations

The first attempts output a result that for each of the twenty type/time categories, PEI behaved best for a small number of large hotspots whereas the PAI metric was maximized by a lot of small hotspots. Hence it shows clearly that one should not attempt to satisfy both metrics simultaneously. In addition, since each of the metrics formed an independent sub-competition, it was deemed better to have a good score for one particular metric than unacceptable results for both. Therefore, for each of the twenty type/time categories, it was decided which metric to focus on in our further work. Moreover, our approach was metric-agnostic. Hence, to choose a metric, we just tossed a coin for each of 20 type/time categories. Four types regular grids were examined: parallelogram grids, triangular grids with 3 vertices at a point, triangular grids with 6 vertices at a point, and hexagonal grids. They were parameterized by cell height, width, translations, rotations and bending. No shape proved noticeably better than other ones. Hence, the ultimate decision was to only use unrotated rectangular grids, parameterized by cell height, width, horizontal and vertical shift. Finally, the number of predicted hotspots was also a hyper-parameter. The model was implemented as Python scripts with NumPy and PyTorch packages used. The computations were performed on the Intel R AI DevCloud infrastructure with Intel R Xeon Scalable Processors, together with optimized distributions of Python and PyTorch. We optimized the grid and our model hyper-parameters for each of 20 type/time categories separately.

5 RESULTS

The results allowed us to conclude that for both the PAI and PEI metrics, we were able to find grids and hotspots with quality competing with predictions obtained by authors of more complicated methods described in the literature [14], [6]. Our approach proved especially effective in categories with the biggest number of crimes committed. Since different competitors submitted different grids, we are unable to compare algorithms for a fixed grid created by particular contestants. Therefore, we cannot judge whether the good performance of our models was an effect of thoroughly scouring potential grids or the power of simplicity of our algorithm for a fixed grid, or perhaps both. At this point we can only claim that our pipeline fulfilled its task.

6 DISCUSSION

The data used and the computation time oriented approach in this study can compete with more sophisticated crime forecasting methods existing in the literature. This result is somewhat surprising. One may conclude that the spatio-temporal distribution of crimes committed is too complicated to be estimated well enough with the use of parametric methods. Or maybe the choice of the proper grid matters much more than it seems. Moreover, There is no reason to claim that the good performance of our algorithm is a one-shot success valid only for Lagos since our model contains no part priory adapted to any particular city. Unfortunately, There is no opportunity to compare the quality of crime forecasts done with use of different methods for the same fixed grid. Such research would shed more light on this field.

The advantage of our algorithm for cases with thousands or more crimes to forecast is an interesting phenomenon. It can be attributed to two possible factors: a specific spatial distribution of crimes or computational simplicity. As the number of events increases, the crimes tend to be spatially distributed more regularly, but with the growing importance of single-point peaks. As stated above, for most statistical parametric methods it may be intractable to cover a distribution containing both a continuous and a discrete part. Comparing the performance of different models for a fixed grid would bear this out. On the other hand, sophisticated algorithms can paradoxically struggle to find the optimal grid and hotspots when presented with large volumes of training data. A time-consuming training procedure for a fixed grid does not allow one to check a sufficient number of potential grids. This problem may be addressed by more efficient algorithms’ implementations and significantly increasing computing resources. Also, adding more constraints on the admissible grid shapes clearly solves the problem, though it also makes it less universal.

Finally, we note that in the perspective of maintaining and updating the crime forecasting system, using only the historical crime data seems to be a good solution. It is hard to find any non-constant external factor which can both influence future crimes and be easier to predict than crimes themselves. Besides, the impact of any hidden important feature is ultimately reflected in the historical data. Moreover, changes in the spatial crime distribution caused by system-driven preventive police activities may be not easy to manage when external data sources are used for forecasting. At the same time, a forecasting system based on merely historical data is able to simply return to the current crime distribution.

Acknowledgments.

We acknowledge the support and sponsorship provided by Covenant University through the Centre for Research, Innovation and Discovery (CUCRID).

REFERENCES

[1] M. A. Boni and M. S. Gerber. Automatic optimization of localized kernel density estimation for hotspot policing. In Proc. 15th IEEE Int. Conf. Machine Learning and Applications (ICMLA), pages 32–38, Dec. 2016. doi: 10.1109/ICMLA.2016.0015.

[2] A. Falade, S. Misra, N. Omoregbe, R. Damasevicius and R. Maskeliunas (2019), An IOT-Based Architecture for Crime Management in Nigeria. Springer Nature Singapore Pte Ltd. 2019 *Data Engineering and Applications*,https://doi.org/10.1007/978-981-13-6351-1_19

[3] S. Chainey, L. Tompson, and S. Uhlig. The utility of hotspot mapping for predicting spatial patterns of crime. Security Journal, 21:4–28, 2008. ISSN 0955-1662. doi: 10.1057/palgrave.sj.8350066.

[4] J. Cohen, W. L. Gorr, and A. M. Olligschlaeger. Leading indicators and spatial interactions: A crime forecasting model for proactive police deployment. Geographical Analysis, 39:105–127, 2007. ISSN 0016-7363. doi: 10.1111/j.1538-4632.2006.00697.x.

[5] M. Fielding and V. Jones. 'disrupting the optimal forager': Predictive risk mapping and domestic burglary reduction in trafford, greater manchester. International Journal of Police Science & Management, 14: 30–41, 2012. ISSN 1461-3557. doi: 10.1350/ijps.2012.14.1.260.

[6] S. Flaxman, M. Chirico, P. Pereira, and C. Loeffler. Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the nij "Real-Time Crime Forecasting Challenge". 2018.

[7] W. Gorr, A. Olligschlaeger, and Y. Thompson. Short-term forecasting of crime. International Journal of Forecasting, 19:579–594, 2003. ISSN 0169-2070. doi: 10.1016/s0169-2070(03)00092-x.

[8] W. L. Gorr and Y. Lee. Early warning system for temporary crime hot spots. Journal of Quantitative Criminology, 31:25–47, 2015. ISSN 0748-4518. doi: 10.1007/s10940-014-9223-8.

[9] J. M. Hunt. Do Crime Hot Spots Move? Exploring the Effects of the Modifiable Areal Unit Problem and Modifiable Temporal Unit Problem on Crime Hot Spot Stability. PhD thesis, 2016.

[10] L. W. Kennedy, J. M. Caplan, and E. Piza. Risk clusters, hotspots, and spatial intelligence: Risk terrain modeling as an algorithm for police resource allocation strategies. Journal of Quantitative Criminology, 27:339–362, 2011. ISSN 0748-4518. doi: 10.1007/s10940-010-9126-2.

[11] H. Liu and D. E. Brown. Criminal incident prediction using a point-pattern-based density model. International Journal of Forecasting, 19:603–622, 2003. ISSN 0169-2070. doi: 10.1016/s0169-2070(03)00094-3.

[12] C. Loeffler and S. Flaxman. Is gun violence contagious? A spatiotemporal test. Journal of Quantitative Criminology, 2017. ISSN 0748-4518. doi: 10.1007/s10940-017-9363-8.

[13] G. Mohler. Marked point process hotspot maps for homicide and gun crime prediction in chicago. International Journal of Forecasting, 30:491–497, 2014. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2014. 01.004.

[14] G. Mohler and M. D. Porter. Rotational grid, PAI-maximizing crime forecasts. to appear in Statistical Analysis and Data Mining, 2018.

[15] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106:100–108, 2011. ISSN 0162-1459. doi: 10.1198/jasa.2011.ap09546.

[16] G. O. Mohler, M. B. Short, S. Malinowski, M. Johnson, G. E. Tita, A. L. Bertozzi, and P. J. rantingham. Randomized controlled field trials of predictive policing. Journal of the American Statistical Association, 110:1399–1411, 2015. ISSN 0162-1459. doi: 10.1080/01621459.2015.1077710.

[17] National Institute of Justice. Real-Time Crime Forecasting Challenge. https://www.nij.gov/funding/Pages/fy16-crime-forecasting-challenge.aspx, 2017. URL https://www.nij.gov/funding/Pages/fy16-crime-forecasting-challenge.aspx.

[18] W. Perry, B. McInnis, C. Price, S. Smith, and J. Hollywood. Predictive policing: The role of crime forecasting in law enforcement operations. Technical report, RAND Corporation, Santa Monica, 2013.

[19] M. D. Porter and B. J. Reich. Evaluating temporally weighted kernel density methods for predicting the next event location in a series. Annals of GIS, 18:225–240, 2012. ISSN 1947-5683. doi: 10.1080/19475683.2012.691904.

[20] L. A. Rastrigin. About convergence of random search method in extremal control of multi-parameter systems. Automation and Remote Control, 24:1467–1473, 1963.

[21] G. Rosser and T. Cheng. Improving the robustness and accuracy of crime prediction with the self-exciting point process through isotropic triggering. Applied Spatial Analysis and Policy, pages 1–21, 2016. ISSN 1874-463X. doi: 10.1007/s12061-016-9198-y.

[22] M. A. Taddy. Autoregressive mixture models for dynamic spatial poisson processes: Application to tracking intensity of violent crime. Journal of the American Statistical Association, 105:1403–1417, 2010. ISSN 0162-1459. doi: 10.1198/jasa.2010.ap09655.

[23] X. Wang, M. S. Gerber, and D. E. Brown. Automatic crime prediction using events extracted from twitter posts. In Social Computing Behavioral - Cultural Modeling and Prediction, pages 231–238. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-29047-3_28.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 3127-3132

[24] K. 190 J. Bowers, S. D. Johnson, and K. Pease. Prospective hot-spotting: The future of crime mapping? The British Journal of Criminology, 44:641–658, 2004. ISSN 0007-0955. doi: 10.1093/bjc/azh036.

[25] L.K. Ajayi, A. A . Azeta, I.T Owolabi, C, F Damilola, A. E Azeta, and O. Amosu. Current Trends in Workflow Mining. Journal of Physics: Conference Series Volume 1299, Issue 1, 7 October 2019.

[26] I. Eweoya, A. Ayodele, A, Azeta, O. Olatunji. Fraud prediction in bank credit administration: A systematic literature review. Journal of Theoretical and Applied Information Technology Open Access Volume 97, Issue 11, 15 June 2019, Pages 3135-3157