Making Crime Pay: Crime Cluster Analysis, Opportunities for Review of Law Enforcement Resource Utilisation

Tom Donoghue*, Ryan Donovan* Badal Kataria* and Dara Adeyemo*
* School of Computing, National College of Ireland, Dublin, Ireland
Email: {x16103491, x16104269, x16110285, x16100379} @student.ncirl.ie

***Abstract*—The amount of data generated related to crime is increasing. The increase is due to new devices and methods of generation, additional points of capture, processing and storage of data at scale. Assisting law enforcement bodies to gain an operational benefit from crime data implies that the output of analysis and data mining should be simple to understand and easy to act upon. An unsupervised learning k-means clustering technique is used to identify and focus on a set of crimes recorded in the city of Chicago. The crime data is supplemented with orthogonal temperature and unemployment data. ANOVA and Kruskal-Wallis statistical tests assess the temporal significance in crimes clusters. The findings indicate various crime hot spots which are temporal and location specific, and therefore may act as input to the scheduling and allocation of policing resources.**

I. Introduction

Public bodies are faced with the perennial issue of allocating a budget for law enforcement services. Their attention is drawn to the benefit of crime prevention measures. The generation and availability of crime data is increasing by the augmentation of mobile, Internet of Things devices and with traditional law enforecement computer systems [1] [2]. Mining the increasing amount of crime data to obtain easily understood and valuable information for decision makers is a challenging task [3]. Research in this area has varied in its use of crime data from how it may assist the rehabilitation of reoccurring offenders [1], to spatial and temporal analysis. This assists in visualising and potentially predicting crime [4][5].

Using the metropolitan crime figures collected in Chicago between 2014 and 2016[1], concentrations of similar crimes, their location and temporal patterns are explored. The crime data is augmented with orthogonal data for unemployment and temperature for the same time period. The extraction of information from the data may assist law enforcement agencies to gain operational efficiencies. Crime disrupts society by its negative impact on individuals, families, community and business [6]. Policing services saddled with the demands of reducing crime whilst encountering severe budget cuts appear to face an insurmountable task. The costs of provision should be balanced with potential benefits and as[1] suggest an upside is obtainable when taking a proactive approach to stopping crime at an early age. Similarly, taking this study, how might such an opportunity benefit present itself for possible input to law enforcement policy decision making? Personnel with crime expertise are capable of producing meaningful information from crime data. These experts may become swamped by the amount crime as the data it generates increases. The use of machine learning may reduce the amount of time and effort required to process the data whilst complementing existing human knowledge [7]. To draw out some early findings, the scope of the crime types are grouped in the investigation. The crimes across 50 council wards are clustered using unsupervised learning k-means clustering. From the 5 clusters produced one cluster contains 2 adjacent western Chicago wards (24 and 28) with the following crimes: prostitution, narcotics, battery, gambling and interference with a public officer. The cluster is labelled as 'vice' based on the crime types. Further examining the data based on day of the week and hour of the day may enable law enforcement bodies to make resource allocation decisions: gain operational performance improvements in areas associated with review of staff rotas and overtime, redeployment of staff better skilled to deal with these crime patterns. This in turn may improve the approach to crime prevention, on the spot activities and crime resolution. Taking a proactive approach on crime prevention and interaction with social and other support services to offer an aligned and holistic set of services. Hot spot policing is not a new approach to crime prevention and [8] confers in their study that the proactive placement of police resources has a positive impact on ameliorating crime figures in 7 out of 9 targeted hot spots. An issue of targeting hot spots is that the crime just transitions elsewhere, but the study indicates that in 5 of the hot spots monitored this was not the case. The following section examines related work in crime data mining, its association in assisting law enforcement agencies to address tight budgets in the allocation of scarce resources. Section 3 covers the approach taken: addressing the business question, pre-processing and initial exploration of the crime data, the trial of different machine learning techniques. Section 4 evaluates the results of the machine leaning models, visualisation of the spatial and temporal patterns explored. Section 5 covers the conclusion and areas for future research.

II. Related Work

A review of the literature related to crime data across the areas of: police resources and utilisation, the application of

[1]https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2 [Accessed April 17, 2017]

machine learning techniques, data visualisation and use of orthogonal data (unemployment and temperature) is covered below. The challenges facing law enforcement and the potential methods which may assist easing these obstacles is also covered.

A. Police Resources & Utilization

Resources available to the police departments in the United States have seen a notable drop since the economic crisis of early 2008. Over 50% of 700 plus police agencies surveyed by the Police Executive Research Forum in 2013 revealed that they were experiencing budget cuts compared to 2010. 40% of those agencies who experienced budget cuts in 2010 further expected an average of 5.3% in budget cuts for 2013 [9]. Although, the situation is improving, it is clear that the police are still under pressure. This has led to cutbacks in police benefits, recruitment and training, and a spike in staff layoffs and retirement incentives [9], [10]. As a result of the reduction in resources, some police departments have become more selective in the crimes they assign resources by focusing more on emergencies over non-emergencies [11]. The preferred method to optimise resources, as [10] concur, is to consider the workload, performance objectives and work schedules, but because this may be expensive to implement, they propose an approach which uses service calls as the main metric for allocating resources [10]. However only about 5% of calls present an opportunity for an officer to make arrests or at least intervene in the situation (cited from a citation in [12], but the original citation wasnt found). This suggests that the volume of service calls is not necessarily the best indicator for resource allocation as calls may exaggerate the need for more personnel deployment. Police knowledge and the information in police computer systems as, [13] and [7] concur, may offer additional combinational strength in the recognition of crime patterns. It is also possible, as [13] find, that both sources of information may lack consistency hence leading to knowledge gaps.

B. Clustering Technique

The pattern of crime often occur in clusters as, [14] suggest, localised both in time, location and types that co-occur. Creating models for predicting crime hot spots and consequently hot spot policing helps in making the best use of limited police resources [6], [5]. Using k-means clustering, it is possible to create new k groups of items which are similar. Instances are identified with a specific cluster based on their distance to closest the centroid in a high dimensional space. The mean position of the centroid is recalculated on each cycle as new instances are associated its change is position to the point when there is no further change [15]. Clustering, a common method in unsupervised data mining has been used to categorize crimes and identify patterns [7], [16]. [6], [17] adopt a systematic k-means clustering method for crime pattern detection by assigning weights to different crime attributes which is an efficient method of detecting crime patterns but requires domain expertise to assign the attribute weights.

C. Association Rules and Decision Trees

Crime patterns have been investigated through mining association rules. Frequent crime patterns were discovered by [17] using apriori algorithms. Fuzzy association rules applied by [18] also show crime patterns that were consistent over multiple regions.They also find some crime association rules which were surprising to the crime experts, further highlighting how useful this technique can be in detecting and understanding crime patterns that may not be immediately obvious to the analyst. [19] go further by finding crime patterns in their research. The authors use a decision tree model, trained by learning from these observed crime patterns for prediction of future crimes. A decision tree is a form of supervised learning where the class label is already known. One of the benefits of a decision tree is that they are easily understood due to their graphical representation and the simple test conducted at each node. The tree is built from a set of rules which partition the data by examining value frequency of attributes. At each node the attribute's value is evaluated and depending on the outcome it either takes a route to the next decision node or terminates in a leaf node. A leaf or terminal node indicates the examples predicted class [20]

D. Visualization

Visual analytics have also been used in crime patterns detection. Graph theory clustering method proposed by [21] employ correlation analysis as a prelude to clustering. Correlations were seen among different crime attributes of different crime types. Their analyses reveal jurisdictions with similar crime patterns, an information which may be useful to policy makers in identifying common fundamental challenges faced by these jurisdictions and how resources can be shared among them. In another method, [22] propose a graphical model that is capable of efficiently extracting patterns from large heterogeneous datasets and show geospatial co-distribution relations among crime incidents, socio-economic, socio-demographic and spatial features. The growing amount of available data makes graphical methods challenging to analyse crime.

E. Seasonality of Crime and Unemployment

Investigations using regression analysis to determine the relationship between crime and factors relating to temperature, time of the day and unemployment are evident in the literature. Crime was found to vary with time of day, day of the week, and season of the year with the occurrence of rape and domestic crimes [23]. A similar trend as, [24], observe a noticeable spike in violence crimes on weekends, during the spring and summer months (i.e May to September), and during periods of darkness in Minneapolis. [25] observe a positive and significant effect of unemployment on some property crimes (burglary, car theft and bike theft) [26] finds that unemployment and all crime rates have a positive correlation in Italian provinces.

III. Methodology

A. Business Question

The CRisp-DM methodology [27] is followed to provide a logical direction for the project. The business question is first postulated in order to direct acquisition of the datasets: "How could machine learning point the Chicago Police Department to better optimise their allocation of resources?" A dataset with temporal and locational crime information is warranted. Also, according to [28], a hotter temperature is correlated with an increase in crime when features like social, economic and demographics are statistically controlled for. Unemployment also has an effect on the increase in crime rates as [25] points out. Although they caution that people that receive unemployment compensation and unregistered employees should be precluded from a crime analysis where possible. Hence, an unemployment data set is also warranted.

B. Pre-Processing

The source data is gathered from three separate sources which is cleansed and integrated into one flat file. Using SQL Server, the source data is loaded into three tables with the attributes irrelevant to this study being disregarded. Next, SQL Server Integration Service is used to upload the data into a fourth table to integrate all the sources. The data is imported to a flat CSV file to be used in the machine learning processes. Highlights of the individual datasets are described below: All the datasets ranged from 2014 to 2016. A crime dataset was selected from[2] which contains crimes committed in the Chicago area. It contained temporal and locational data at minute level granularity. It also had different crime types and district attributes. A temperature dataset was gathered[3] which has the daily average temperatures for the city of Chicago. The unemployment dataset[4] gathered had the average monthly unemployment rate. There were some challenges in the overall pre-processing phase. The granularity in the crimes was at minute level but the average temperature and unemployment datasets were at daily and monthly levels respectively. Also, the unemployment rate included people receiving unemployment benefits and unregistered workers. Using unemployment could impact the results as pointed out by [25]. There are missing values in the data that need to be fixed. Seven days were missing temperature data, so the average of 2 weeks prior and after were used to fill in place. Some values needed to be unified due to different value referring to the same category of crime. For example, there were 3 crimes reported as "NON-CRIMINAL", "Non_Criminal" and "NON CRIMINAL" that were unified into one value "NonCriminal". The final dataset contains 803,082 rows for the 3 year period which is explored by graphing, preliminary machine learning techniques and statistically analysed (i.e. mean and standard deviation) in RapidMiner. There were many iterations of the data where binary variables were added. There were also aggregations (e.g. aggregation of days and hours).

[2]https://data.cityofchicago.org/view/5cd6-ry5g [Accessed April 17, 2017]

[3]https://www.glerl.noaa.gov//metdata/chi/archive [Accessed April 17, 2017]

[4]http://www.ides.illinois.gov/LMI/Local Area Unemployment Statistics LAUS/ILChicagoMetroAreaUnemploymentRates/ILlaus-seasadj.PDF [Accessed April 17, 2017]

C. Machine Learning Algorithms

An initial trial using the decision tree machine learning model was used on the early version of the cleansed dataset. The machine learning application RapidMiner is used because of its graphical user interface which makes it is simple to configure a workflow. A workflow caters for a series of operators which can be dragged into place to perform a specific process. Cross validation was used to test the performance of the decision tree algorithm against the set of 803,082 instances. The model is trained and tested using a stratified sampling (which builds random subsets for training and testing which reflect the same proportion of the class label as contained in the original example set) with the 10 folds (number of subsets created, each one containing equal numbers of instances from the original example dataset). The resulting decision tree produced over 380 nodes which is far too complex. One recognised method of overcoming this issue is to bin the attributes into ranges and use these ranges to train and test the model [29]. This approach is taken to bin temperature, unemployment, and temporal attributes. 4 temperature zones are created: cold ($< 9^o$C), mild (9^o to 20^oC), warm (20^o to 29^oC), and hot ($> 29^o$C). Three unemployment zones were made: below 6%, 6 to 8% and above 8%. Time zones for each hour was created along with 4 zones at 6-hour intervals starting at 00:00 (early morning, morning, afternoon and evening). A further decision tree was trained and tested, but continued to produce unsatisfactory results.

A decision was made to explore clustering using both unsupervised and supervised learning approaches; k-means and using kNN respectively. kNN is conducted to predict the counts of arrests for a given day. Knowing the likely amount of arrests for particular crimes for any given day might be useful for the police force in allocating officer resources for these days. The crime types were separated into binary variables before aggregating to the daily level in order to get counts for each crime on a given day. The data was split into a training set of 80% and the remaining 20% reserved for testing. kNN requires a categorical variable for the class label, arrest counts were separated into four levels by using min-max normalization. The confusion matrix was gathered of which is discussed in the results section. k-means clustering is used to examine clusters of crime in Chicago wards and also for crimes that occurred on Fridays. For the latter, data was aggregated to include Fridays. The input examples were from the 50 Chicago wards with the counts of the 30 crime categories reported. As the crime count range differed, the data was normalised to standardise the unit of scale (enabling the comparison of like with like) before initiating the algorithm. The value for k was trialled between 3 and 10. k=5 appeared to provide the best results with a suitable number of clusters. One drawback of the k-means as [15] concurs is the inability to suggest the correct number of cluster to choose at the

beginning. The clusters produced may be subject to qualitative evaluation using domain expertise. Focusing on the clustering outcome; cluster 3 contains the higher mean scores. The cluster is named the "vice" cluster and contains prostitution, narcotics, battery, gambling and interference with a public officer with high concentration of these crimes in wards, 24 and 28 of Western Chicago. The focus of the exploration is directed towards the vice cluster.

In investigating the temporal patterns of these crimes, a question arises; could there be a significant difference in crime frequencies depending on the day of the week (e.g. are batteries less prevalent on Thursdays than Saturdays)? A one way ANOVA provides a suitable statistical test to provide an indication of significant difference between groups. There appears to be differences in the amount of crimes for different days of the week in initial investigation. One-way ANOVA and Kruskal-Wallis tests were conducted to statistically prove if there is a significant difference in the frequency of crime occurrence on different days for each crime at a 95% confidence level. The hypothesis test is:

> *H_O: there is no difference between the means of the days for rates of battery crime reported*
> *H_A: there is a difference between at least one of the means for the rates of battery crime reported*

The association rules method is used to investigate interesting patterns in the data by using Apriori algorithm. This is conducted with the "arules" package in R. The 5 crimes from the 2 wards chosen in k-means cluster analysis is then aggregated temporally into monthly, daily, and time of day zones. Temperature zones, unemployment zones, location of crimes (e.g. alley, street, residence etc.) and "Arrested or Not" attributes are also used. As association rules is an unsupervised technique, it is unknown what patterns could surface from the algorithm. Different support and confidence values were tried and ordered by lift. Both are then adjusted depending on whether there are too few or too many rules being produced. As some rules had implications (e.g. August -> Hot) filtering was applied to only allow crime attributes to be on the right hand side of the implication.

In a similar approach to the kNN, a regression analysis is conducted in SPSS to determine if the number arrests could be predicted for each crime on any given day in the vice cluster. This could help to in the allocation of resources towards combating crime, as explained in kNN, or to predict the amount of a certain crime depending on input variables to schedule specialists in one of the 5 crimes in those 2 wards. Aggregation of the crime data is applied to the daily level and the correlations between all variables is compared. Correlations (Pearsons R) are observed to be below 0.4 and thus not selected for the linear regression. Because, assumptions for normality, non-collinearity between independent, linearity with dependent, residuals following a normal distribution and being independent all pass, a multiple regression to obtain a model equation is conducted next. The regression analysis is conducted in SPSS.

IV. Evaluation / Results

The approach to address the business question of how might mining crime data present actionable information produced the following results. Discoverable patterns are required to indicate where patterns The identification of groups based on the crimes reported of crime of the application of the approach

A. k-means

When using k-means at a value of k=3 to explore Friday crime clusters, the following crimes appeared together: battery, gambling, liquor law violations, narcotics, prostitution, public peace violation and weapons violation. A separate analysis of clustering amongst the 50 wards in Chicago showed similar crimes clustering when using k=5. The following 5 crime appeared with the highest centroid mean values in cluster 3: prostitution 4.57, narcotics 4.19, battery 2.6, gambling 3.63 and interference with a public officer 3.13. The vice cluster also had 2 wards (n=2) in the analysis, namely ward 24 and ward 28. The algorithm was rerun multiple times with different random seeds and produced the same 2 wards. Figure 1 shows a depiction of the clusters and the subset of cluster 3 that was chosen. The results are built using a subsets of date using the vice cluster wards and maintaining focus on the five crimes just named.

B. Temporal Analysis and Association Rules

With this new subset, the temporal distribution of crimes was analyzed. Figure 2 shows the distribution of crimes in the 2 wards for all crimes for each day of the week. Each day followed a similar pattern where crimes would peak at midnight and then go down to the lowest amount at 6AM. Then the crime amounts would go up till midnight again. This could possibly be used by police to schedule officers according to the amounts of crimes happening at certain times and days.

The five crimes were then focused on for the crime distribution over the week. There was a significant effect of the day of the week on the level of battery crime reported, $F(6, 1089) = 12.51$, $p < .001$. Post Hoc comparisons using Tukey HSD indicated that the mean for Sunday (M = 13.65, SD = 4.82), Monday (M = 13.14, SD = 4.486), Tuesday (M = 13.54, SD = 4.152), Wednesday (M = 13.19, SD = 4.075) and Thursday (M = 15.94, SD = 46.9) do not differ significantly from each other. Friday (M = 15.94, SD = 5.099) and Saturday (M = 16.46, SD = 5.244) are significantly different from Sunday, Monday, Tuesday and Wednesday. As a result the null hypothesis is rejected. Figure 3 shows the min-max normalization of crimes for each day of the week. As prostitution, narcotics and gambling violated the homogeneity of variance assumption a Kruskal-Wallis (a non-parametric test which caters for a non-normal distribution) was conducted. The result of the Kruskal-Wallis test indicated, at an alpha value = .05, that prostitution crime is affected by the day of the week, $H(6) = 123.975$, p

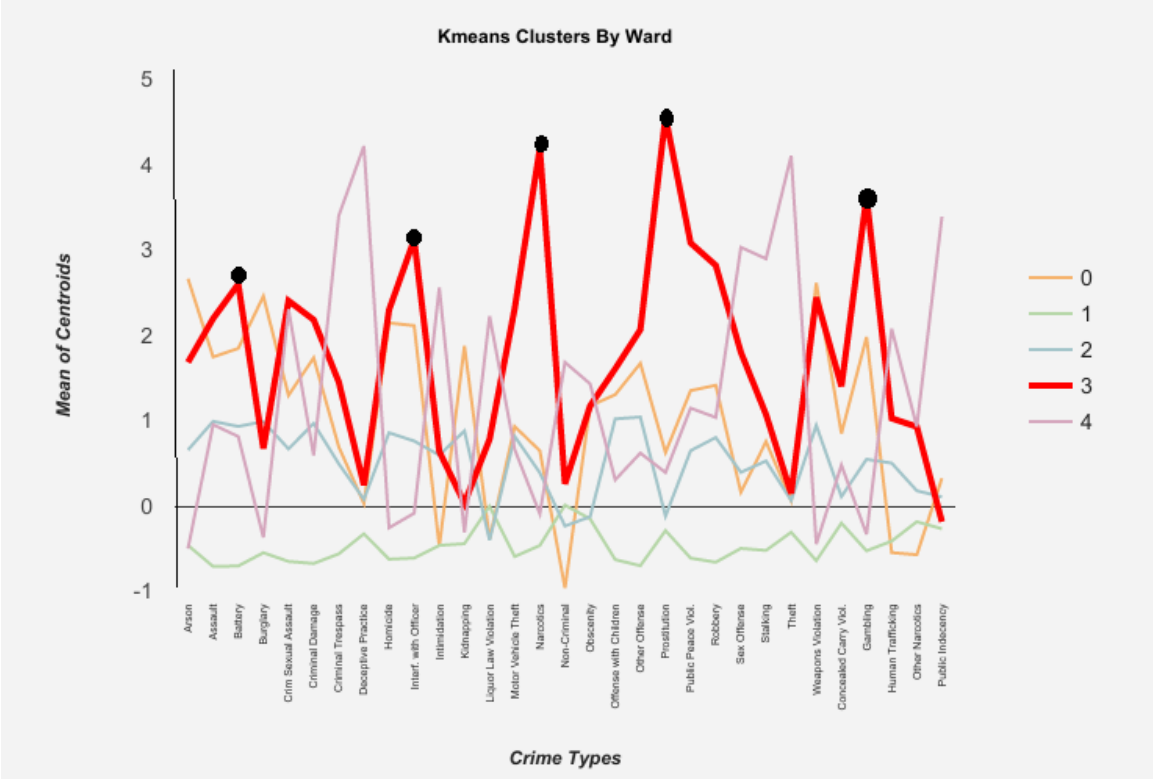

Fig. 1. Shows the 5 crimes chosen by in 2 wards in cluster 3 labeled by black dots

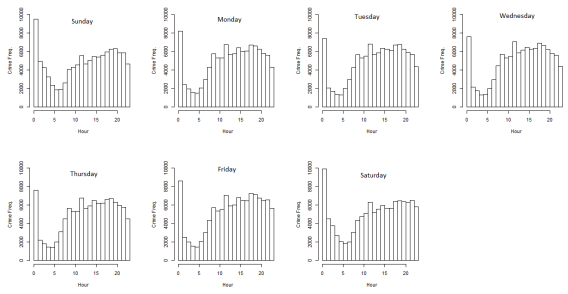

Fig. 2. Shows the distributions of all crimes per hour using 2014-2016 data

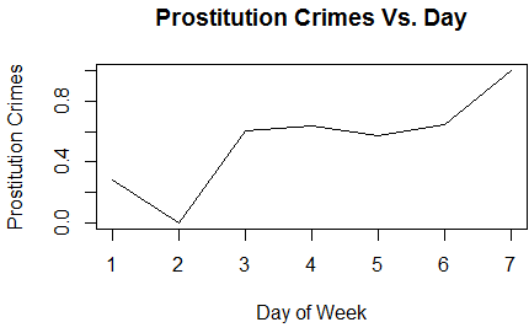

Fig. 4. Prostitution Crimes over 7 days of the week. 1 = Sunday to 7 = Saturday

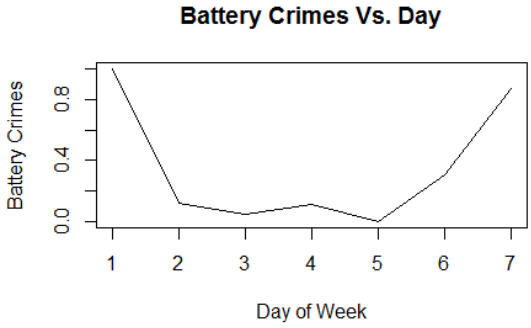

Fig. 3. Battery Crimes over 7 days of the week. 1 = Sunday to 7 = Saturday

$<$ .001 as a result, the null hypothesis was rejected. Figure 4 shows prostitution crimes for the seven days of the week.

Gambling and narcotics also showed a difference in the count for each day as well. The result of the Kruskal-Wallis test indicated, at an alpha value = .05, that gambling is not affected by the day of the week, H(6) = 7.395, p = .286. So the null hypothesis was rejected. The result of the Kruskal-Wallis test indicated, at an alpha value = .05, showed that narcotics count is affected by the day of the week, H(6) = 30.387, $p <$ .001 as a result we reject the null hypothesis.

Investigation was also done for the distributions of crime

counts for batteries, narcotics, and prostitution for the hour and day of the week. as shown in Figure 5, Figure 6 and Figure 7 All the counts of crimes were min-max normalized again. Figure 5 shows that between the hours of midnight to 7AM there are more battery crimes on Saturday and Sundays as compared to the rest of the week. ANOVAs comparing the different days of the week from Midnight to 7AM showed there was a statistically different similarly to previous ANOVAs. From 7AM to midnight there was similar amount of battery crimes for each day of the week. Figure 6 shows the distributions of crimes of narcotics over the hour for 7 days of the week. Every day seemed similar except for the times 5AM to 9AM. An ANOVA was also conducted to show that there was a statistical difference between days for narcotics crimes. Figure 7 shows the similar graph for Prostitution crime, but there was much more variability as shown in the figure. Tuesday to Friday had similar distributions but the Monday did not act similar. Sundays had a high peak in prostitution crime at 2AM but was relatively low the rest of the hours. Future work might investigate the prostitution distribution to maybe learn about why this crime behaves differently than the other 2 crimes just discussed.

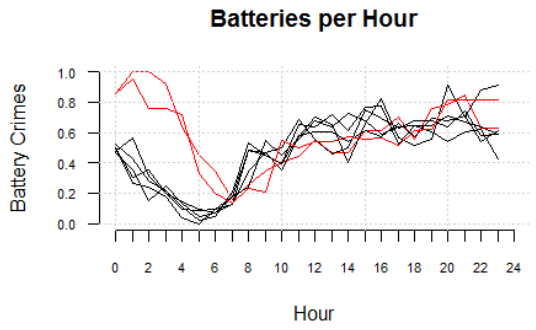

Fig. 5. Compares the distribution of Battery crimes during weekdays and weekends by hour. Red is Saturday and Sunday and Black Monday to Friday

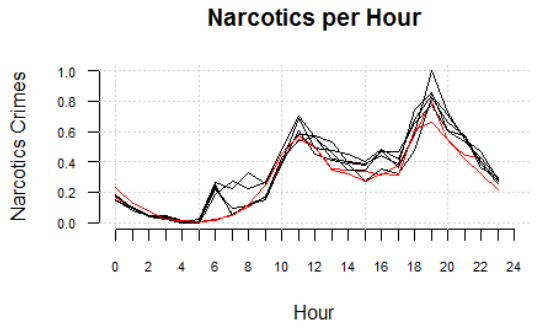

Fig. 6. Compares the distribution of Narcotics crimes during weekdays and weekends by hour. Red is Saturday and Sunday and Black Monday to Friday

Association Rules Mining was run on the ward subset to see if there were any interesting patterns of crime happening

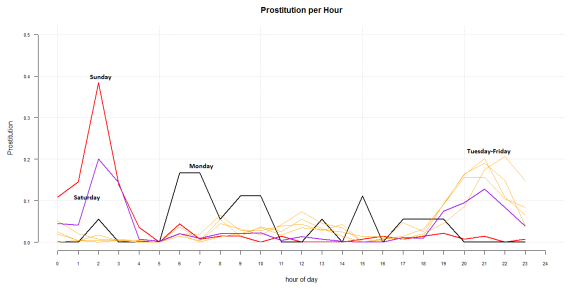

Fig. 7. Compares the distribution of Prostitution crimes over various day by hour. Purple = Saturday, Red = Sunday, Monday = Black and Tuesday-Friday = Orange

in relation to time. A Support of 0.1 and confidence of 0.25 was first tried and ordered by lift. There were only 13 rules found but these were too simple to be of any value (e.g. June indicating hot temperatures or that narcotics indicated an arrest). This was uninteresting because dividing all the narcotic crimes by arrests revealed a 99.7% chance of arrest. The full data set was similar with 99.7% chance of arrests. Filtering was done along with varying the support and confidence but no interesting rules were found. Though association rules failed to produce promising results, it did lead to an investigation of the difference of narcotics arrest for the 3 years. Figure 8 shows the percentage of arrests for narcotic crime over the 3 years. An ANOVA showed that there was a difference between the amount of arrests for the 3 groups (df(1,34), F = 7.912, p = 0.0081 $<$ 0.05). 2015 was similar to 2016 with 95% confidence (df(1,22), F = 0.312, p = 0.582 $>$ 0.05) but 2014 was different than 2015 and 2016 with 95%(df(1,22), F = 9.338, p = 0.00579 $>$ 0.05) and (df(1,22), F = 9.26, p = 0.00597 $>$ 0.05) respectively. Since the mean of 0.995 of percentage of arrests was less than 2015 and 2016, it can be said that narcotics arrests were less than the other 2 years.

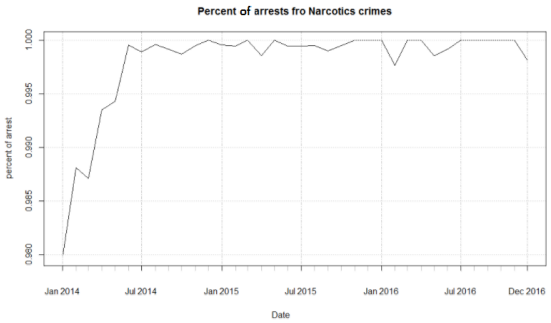

Fig. 8. Shows the percentage of arrests made for Narcotics crimes over 3 years

From the findings, there was a difference in the number of arrests for the given years. Both the total crime and arrests were examined over the 3 years across all wards. The distribution of crime counts for each month over 3 years in Chicago is shown in Figure 9. 2014 made up 35% of the

crimes, 2015 made up 33%, and 2016 had 32%. There is a decline in total reported crimes for each year and as [5] points out this may be due to Chicago police department having decided to focus less on petty crimes (like low-level narcotics crimes) and focus on gun crime and community relations. With this knowledge a graph of the narcotics crimes for all wards for three years was made as shown in Figure 10. The amount of narcotics crimes more than halved from 2014 (28,963) to 2016 (12,421) that there were less narcotics arrests from 2014 (28,963) to 2016 (12,421).

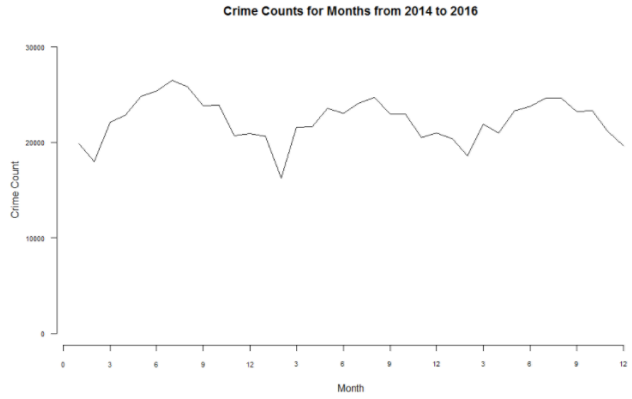

Fig. 9. Shows distribution of crimes from 2014 to 2016 by month

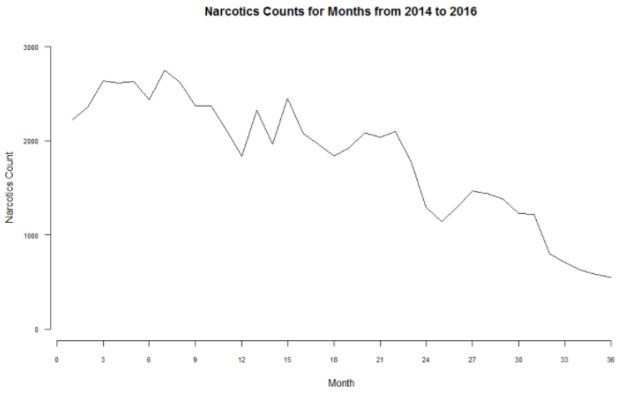

Fig. 10. Shows decline of narcotics crimes from 2014 to 2016 by month

C. Multiple Regression

A multiple regression was done to predict the amount of arrests in the 2 wards with input values of temperature, unemployment, day of week. There was a correlation of R = 0.447 between temperature and battery crimes. The first multiple linear regression to predict arrests only explained 20% of variance (adjusted R = 0.199) and df(1,1094), F = 272.553, p = 0.0). A correlation of R = 0.876 between narcotics counts and arrest counts which was found in association rules analysis.

[5]http://chicago.suntimes.com/news/the-watchdogs-arrests-down-25-percent-in-chicago-this-year/ [Accessed April 17, 2017]

D. Decision Trees

A decision tree analysis was carried out on original data set for all wards and crimes to see if a an arrest could be predicted using day of the of day, temperature, unemployment and crime type. Over 380 nodes were created. This was far too complex to be usable as a predictor for arrest. The main probable cause for this was down to high levels of entropy (the attribute contains a high frequencies of unique values) in the attributes used. One method of overcoming this issue would be to bin the attributes into ranges and use these ranges to train and test the model.

E. kNN

The kNN analysis showed less than promising results. The overall prediction accuracy was 77% and the precisions ranged from 67% to 100% as shown in the confusion matrix in Figure 11. Although, the recall values were ranged from 11% to 88%. For example, the class label 0.75 to 1.00 only had 1 of 9 predicted correctly and for the label 0.50 to 0.75 only 2 of 16 were predicted correctly. Maybe this was because the arrests went down each year

	agg_test_pred				
agg_test_labels	<.25	[0.25,0.5)	[0.5,0.75)	>0.75	Row Total
<.25	67	15	0	0	82
	0.817	0.183	0.000	0.000	0.374
	0.779	0.116	0.000	0.000	
	0.306	0.068	0.000	0.000	
[0.25,0.5)	12	99	1	0	112
	0.107	0.884	0.009	0.000	0.511
	0.140	0.767	0.333	0.000	
	0.055	0.452	0.005	0.000	
[0.5,0.75)	0	14	2	0	16
	0.000	0.875	0.125	0.000	0.073
	0.000	0.109	0.667	0.000	
	0.000	0.064	0.009	0.000	
>0.75	7	1	0	1	9
	0.778	0.111	0.000	0.111	0.041
	0.081	0.008	0.000	1.000	
	0.032	0.005	0.000	0.005	
Column Total	86	129	3	1	219
	0.393	0.589	0.014	0.005	

Fig. 11. Confusion matrix for kNN procedure

V. Conclusion

Conclusions and future work: summarise your findings, and discuss limitations / extensions that were you to have more time, you would do next to improve / extend your study. Summarise the (partial) answer to the research question(s) at a high level, and note the contribution to knowledge the paper has made.

References

[1] W. G. Cohen, Mark A Piquero, Alex R Jennings, "Studying the costs of crime across offender trajectories," *Criminology & Public Policy*, vol. 9, no. 2, pp. 279–305, 2010.

[2] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," *Computer*, vol. 37, no. 4, pp. 50–56, 2004.

[3] C.-H. Yu, M. W. Ward, M. Morabito, and W. Ding, "Crime Forecasting Using Data Mining Techniques," *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 779–786, 2011.

[4] L. Tompson and M. Townsley, "(Looking) Back to the Future: using spacetime patterns to better predict the location of street crime," *International Journal of Police Science & Management*, vol. 12, no. 1, pp. 23–40, 2010.

[5] N. Lazzati and A. A. Menichini, "Hot spot policing: A study of place-based strategies for crime prevention," *Southern Economic Journal*, vol. 82, no. 3, pp. 893–913, 2016.

[6] S. V. Nath, "Crime pattern detection using data mining," in *2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops*, Dec 2006, pp. 41–44.

[7] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," *IEEE Computer Society*, vol. 37, no. 4, pp. 50–56, 2004.

[8] A. A. Braga, *Police Enforcement Strategies to Prevent Crime in Hot Spot Areas*. United States Department of Justice, 2008, vol. 2, no. 2.

[9] C. Wexler, *Policing and the Economic Downturn: Striving for Efficiency Is the New Normal*, 2013.

[10] J. M. Wilson and A. Weiss, "A performance-based approach to police staffing and allocation," *US Department of Justice Office of Community Oriented Policing Services*, 2012.

[11] M. J. Parlow, E. Hall, and R. E. P. A. N. O, "The Great Recession and Its Implications for Community Policing," *Georgia State Univeristy Law Review*, vol. 1, pp. 1–30, 2012.

[12] J. E. Eck and W. Spelman, "Problem-solving: Problem-oriented policing in newport news," 1987.

[13] R. Haining and J. Law, "Combining police perceptions with police records of serious crime areas: a modelling approach." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 170, no. 4, pp. 1019 – 1034, 2007.

[14] J. Zipkin, M. B. Short, and A. Bertozzi, "Cops on the dots in a mathematical model of urban crime and police response," *Dcds-B*, vol. 19, pp. 1479–1506, 2014.

[15] A. Kigerl, "Cyber Crime Nation Typologies : K-Means Clustering of Countries Based on Cyber Crime Rates," vol. 10, no. 2, pp. 147–169, 2016.

[16] U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns," *2016 Second Asian Conference on Defence Technology (ACDT)*, pp. 123–128, 2016.

[17] J. D. E. Sandig, R. M. Somoba, M. B. Concepcion, and B. D. Gerardo, "Mining online gis for crime rate and models based on frequent pattern analysis," vol. 2, 2013, pp. 23–27.

[18] A. L. Buczak and C. M. Gifford, "Fuzzy association rule mining for community crime pattern discovery," *ACM SIGKDD Workshop on Intelligence and Security Informatics - ISI-KDD '10*, pp. 1–10, 2010.

[19] S. Sathyadevan, M. Devan, and S. Surya Gangadharan, "Crime analysis and prediction using data mining," pp. 406–412, 2014.

[20] S. B. Kotsiantis, "Decision trees: A recent overview," *Artificial Intelligence Review*, vol. 39, no. 4, pp. 261–283, 2013.

[21] P. Zhao, M. Darrah, J. Nolan, and C.-Q. Zhang, "Analyses of crime patterns in nibrs data based on a novel graph theory clustering method: Virginia as a case study," *The Scientific World Journal*, 2014.

[22] P. Phillips and I. Lee, "Mining co-distribution patterns for large crime datasets," *Expert Systems with Applications*, vol. 39, no. 14, pp. 11 556–11 563, 2012.

[23] E. G. Cohn, "The prediction of police calls for service: The influence of weather and temporal variables on rape and domestic violence," *Journal of Environmental Psychology*, vol. 13, no. 1, pp. 71–83, 1993.

[24] J. Rotton and E. G. Cohn, "Violence is a curvilinear function of temperature in dallas: a replication." *Journal of personality and social psychology*, vol. 78, no. 6, p. 1074, 2000.

[25] R. Yildiz, O. Ocal, and E. Yildirim, "The effects of unemployment, income and education on crime: Evidence from individual data," *International Journal of Economic Perspectives*, vol. 7, no. 2, p. 32, 2013.

[26] N. Speziale, "Does unemployment increase crime? Evidence from Italian provinces," *Applied Economics Letters*, vol. 21, no. 15, pp. 1083–1089, 2014.

[27] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "Crisp-dm 1.0 step-by-step data mining guide," 2000.

[28] C. A. Anderson, "Temperature and aggression: effects on quarterly, yearly, and city rates of violent and nonviolent crime." *Journal of personality and social psychology*, vol. 52, no. 6, p. 1161, 1987.

[29] F. Berzal, J.-C. Cubero, N. Marn, and D. Snchez, "Building multi-way decision trees with numerical attributes," *Information Sciences*, vol. 165, no. 12, pp. 73 – 90, 2004.