

Crime Prediction Using Regression and Resources Optimization

Bruno Cavadas^{1,2}, Paula Branco^{3,4}, and Sérgio Pereira⁵

¹ Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal

² Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Portugal

³ LIAAD - INESC TEC

⁴ DCC - Faculdade de Ciências - Universidade do Porto

⁵ ALGORITMI Centre, University of Minho, Portugal

brunomcavadas@gmail.com paobranco@gmail.com pereirasrm@gmail.com

Abstract. Violent crime is a well known social problem affecting both the quality of life and the economical development of a society. Its prediction is therefore an important asset for law enforcement agencies, since due to budget constraints, the optimization of resources is of extreme importance. In this work, we tackle both aspects: prediction and optimization.

We propose to predict violent crime using regression and optimize the distribution of police officers through an Integer Linear Programming formulation, taking into account the previous predictions. Although some of the optimization data are synthetic, we propose it as a possible approach for the problem. Experiments showed that Random Forest performs better among the other evaluated learners, after applying the SmoteR algorithm to cope with the rare extreme values. The most severe violent crime rates were predicted for southern states, in accordance with state reports. Accordingly, these were the states with more police officers assigned during optimization.

Keywords: violent crime, prediction, smoteR, regression, optimization

1 Introduction

Violent crime is a severe problem in society. Its prediction can be useful for the law enforcement agents to identify problematic regions to patrol. Additionally, it can be a valuable information to optimize available resources ahead of time.

In the United States of America (USA), according to the Uniform Crime Reports (UCR) published by the Federal Bureau of Investigation (FBI) [1], violent crimes imply the use of force or threat of using force, such as rape, murder, robbery, aggravated assault, and non-negligent manslaughter. In 2013, it was reported 1,163,146 violent crimes, with an average of 367.9 per 100k inhabitants. This was equivalent to one violent crime every 27.1 seconds. In 2012, according to the United States Department of Labor [2], there were 780,000 police officers

and detectives in the USA, with a median salary of \$56,980 per year. Therefore, the optimization of police officers can be useful to optimize costs, while guaranteeing the safety of the population.

In this paper, the contributions are twofold. Firstly, we propose to predict the violent crime per 100k population using regression. To the best of our knowledge, this is the first time that such problem is tackled in this way. Moreover, we pre-process the data using smoteR algorithm to improve predictions on the most critical values: the extreme high. Having the predictions, we also propose an Integer Linear Programming formulation for the optimization of police officers distribution across states. This distribution takes into account the crime severity, population, density and budget of the states.

The remaining of the paper is organized as follows. In Section 2 a brief survey on related work is presented. Materials and methods are exposed in Section 3, including the description of the data set, the prediction-related procedures and the optimization scheme. Then, in Section 4, results are presented and discussed, while in Section 5 the main conclusions are pointed out.

2 Related work

Crime prediction has been extensively studied throughout the literature due to its relevance to society. These studies employ diverse machine learning techniques to tackle the crime forecasting problem.

Nath [3] combined K-means clustering and a weighting algorithm, considering a geographical approach, for the clustering of crimes according to their types. Liu et al. [4] proposed a search engine for extracting, indexing, querying and visualizing crime information using spatial, temporal, and textual information and a scoring system to rank the data. Shah et al. [5] went a step further and proposed CROWDSAFE for real-time and location-based crime incident searching and reporting, taking into account Internet crowd sourcing and portable smart devices. Automatic crime prediction events based on the extraction of Twitter posts has also been reported [6].

Regarding the UCI data set used in this work, Iqbal et al. [7] compared Naive Bayesian and decision trees methods by dividing the data set into three classes based on the risk level (Low, medium and high). In this study, decision trees outperformed Naive Bayesian algorithms, but the pre-processing procedures were rudimentary. Somayeh Shojaee et al. [8], applied a more rigorous data processing methodology for a binary class and applied the usage of two different feature selection methods to a wider range of learning algorithms (Naive Bayesian, decision trees, support vector machine, neural networks and K-Nearest neighbors). In these studies no class balancing methodologies were employed. Other approaches such as the fuzzy association rule mining [9] and case-based editing [10] have also been performed.

After prediction, optimization of resources can be achieved by several strategies. Donovan et al. [11] used integer linear programming for the optimization of fire-fighting resources, solving one of the most commonly constrains faced by fire

managers. The same strategy was used by Caulkins et al. [12] in the optimization of software system security measures given a fixed budget.

Regarding the problem of police officer optimization, Mitchell [13] used a P-median model to determine the patrol areas in California, while Daskin [14] applied a Backup Coverage Model to maximize the number of areas covered. More recently, Li et al. [15] relied on the concept of “crime hot-spots” to create a cross entropy approach to produce randomized optimal patrol routes.

3 Materials and Methods

3.1 Data Set Description

The *Communities and Crime Unnormalized Data Set*¹ provides information on several crimes in the USA, combining socio-economical and law enforcement data from 90’ Census, 1990 Law Enforcement Management and Admin Stats survey and the 1995 FBI UCR. It includes 2215 examples, 124 numeric and 1 nominal attribute. It also contains 4 non-predictive attributes with information about the community name, county, code and fold. Among the several possible target variables we chose the number of violent crimes per 100k population.

3.2 Prediction

We started by pre-processing the data set. The violent crime is our target variable, thus we removed all the other 17 possible target variables contained in the data set. We also eliminated all the examples that had a missing value on our target variable and removed all the attributes that had more than 80% of missing values. The data set contained four non-predictive attributes, which we have also eliminated. Finally, we have removed one more example that still had a missing value, and have normalized all the remaining attributes.

Although this problem was previously tackled as a classification task, we opted for addressing it as a regression task. This is an innovative aspect of our proposal and this choice is also based on the fact that we will use the numeric results obtained with the predictions for solving an optimization problem. Therefore, it makes sense to use a continuous variable throughout the work, instead of discretizing the target variable and latter recovering a numeric value.

Another challenge involving this data set is the high number of attributes. To address this problem we have applied the same feature selection scheme with two different percentages. The scheme applies a hierarchical clustering analysis, using the Pearson Correlation Coefficient. This step removes a percentage of the features less correlated with the target variable. Then, a Random Forest (RF) learner is applied to compute the remaining features importance based on the impact in the Mean Squared Error. A percentage of the most important features provided is selected. Two different sets of features were selected by applying

¹ available at UCI repository in <https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized>.

different percentages in the previous scheme. In one of the pre-processed data we aimed at obtaining 50% of the original features and in the other the goal was to select only 30% of the original features. This way we obtained two data sets with 52 and 32 features corresponding to 50% and 30% percentages.

In our regression problem we are interested in predicting the number of violent crimes per 100k inhabitants. However, we are more concerned with the errors made in the higher values of the target variable, i.e., the consequences of missing a high value of violent crimes by predicting it as low are worst than the reverse type of error. The extreme high values of the violent crime variable are the most important and yet the less represented in the data set. When addressed as a classification problem, this is clearly a problem with imbalanced classes, where the most important class has few examples. SmoteR algorithm is a proposal to address this type of problems within regression which was presented in [16,17]. This proposal uses the notion of utility-based Regression [18] and relies on the definition of a relevance function. The relevance function expresses the user preferences regarding the importance assigned to the target variable range. Ribeiro [18] proposes automatic methods for estimating the relevance function of the target variable. We have used those methods because they correspond to our specific concerns: the extreme rare values are the most important. The essential idea of SmoteR algorithm is to balance the data set by under-sampling the most frequent cases and over-sampling the rare extreme examples. The over-sampling strategy generates new synthetic examples by interpolating existing rare cases. More details can be obtained in [16,17]. The motivation for applying this procedure is to force the learning systems to focus on the rare extreme cases which would be difficult to achieve in the original imbalanced data. Our experiences included several variants of smoteR which were applied to the two pre-processed data sets. The smoteR variants used in the experiences included all combinations of the following parameters: under-sampling percentage 50% and 100%; over-sampling percentage 200% and 400%; number of neighbours 5.

For the prediction task we have used three learning algorithms: Support Vector Machines (SVM), RF and Multivariate Adaptive Regression Splines (MARS). More details on the experimented parameters and the evaluation are described in Section 4.1.

3.3 Optimization through Integer Linear Programming

Given the predicted violent crime per 100k population, we propose to optimize the distribution of available police officers by state. We present our proposal as a proof of concept, since more detailed data and insight into the problem would be needed to implement a more realistic solution. Given that the number of officers by state is an integer quantity, it is used Integer Linear Programming. To solve the optimization problem it was applied the Branch-and-bound algorithm.

Problem formulation We considered as resources a certain amount of police officers to freely distribute by the states of the USA. The optimization takes

into account the predictions on violent crime per 100k population to assign more officers by the states with more violent criminality. This assignment is constrained by an ideal number of officers that each state would like to receive and the available budget. However, every state should receive a minimum amount of officers to guarantee the security of its citizens.

In the data set, the instances are defined by communities, with several of them belonging to the same state. Since we wanted to distribute officers by state, it was calculated the mean violent crime predictions by state.

The optimization problem was defined as,

$$\begin{aligned}
& \text{maximize} && \sum_{i=1}^m s_i x_i \\
& \text{subject to} && \sum_{i=1}^m x_i = N; \quad x_i \leq H_i; \\
& && x_i \geq f_i H_i; \quad c_i x_i \leq B_i; \\
& && x_i \in \mathbb{N}
\end{aligned}$$

where $i \in \{1, \dots, m\}$ indexes each of the m states, with $m = 46$, x_i is the number of officers to distribute by state, s_i is the violent crime predictions by state, H_i is the ideal number of officers by state, f_i is a fraction on the ideal number of officers that each state accepts as the minimum, c_i is the cost that each state should pay for each officer, and B_i is the available budget for each state.

The ideal number of officers was defined in function of the violent crime prediction of the state and the population (number of citizens), since bigger populations, with more violent crime, have higher demands regarding police officers. To this end, the violent crime predictions were scaled (s_{s_i}) to the interval $[v_l, v_h]$. This way, it acts as a proportion on the population. However, since some populations have millions of citizens, this value was divided by 100 to get more realistic estimates for the ideal number of officers. So,

$$H_i = \frac{s_{s_i} p_i}{100} \tag{1}$$

where p_i is the real population of the state i .

It was defined that the minimum number of officers should be a fraction on the ideal number, taking into account the crime predictions. Defining a lower (l_b) and an upper (u_b) bound for the fraction, the previously scaled violent crime predictions are linearly mapped to the interval $[l_b, u_b]$. Knowing that it is in the interval $[v_l, v_h]$, the fraction on the ideal number of officers is calculated as,

$$f_i = \frac{s_i - v_l}{v_h - v_l} (u_b - l_b) + l_b \tag{2}$$

Budget was defined in function of the population and its density. Such definition is based on the intuition that a small and less dense population needs less budget and officers than a highly dense and big population. However, the population numbers are several orders of magnitude higher than density, which would make the effect of density negligible. So, we have rescaled both population

and density to the range $[0, 100]$ (p_{s_i} and d_{s_i}). Moreover, the budget for each state is a part of the total national budget (B_T). So, B_i was calculated as

$$B_i = \frac{(d_{s_i} + a \cdot p_{s_i}) B_T}{\sum_{i=1}^m d_{s_i} + a \cdot p_{s_i}} \quad (3)$$

where $a > 0$ is a parameter to tune the weight of the density and population over the budget calculation.

4 Experimental Analysis

We have divided our problem, and analysis, into two sub-problems: prediction and optimization. In this section, we describe the tools, metrics, and evaluation methodology for each sub-problem. Then we focus in each sub-problem results.

4.1 Experimental Setup

Prediction The main goal of our experiments is to select one of the two pre-processed data sets, a smoteR variant (in case it has a positive impact) and a model (among SVM, RF and MARS) to apply in the optimization task.

The experiments were conducted with R software. Table 1 summarizes the learning algorithms that were used and the respective parameter variants. All combinations of parameters were tried for the learning algorithms, which led to 4 SVM variants, 6 RF variants and 8 MARS variants.

We started by splitting each data set in train and test sets, approximately corresponding to 80% and 20% of the data. The test set was held apart to be used in the optimization, after predicting its crime severities. This set was randomly built with stratification and with the condition of including at least one example for each possible state of the USA.

In imbalanced domains, it is necessary to use adequate metrics since traditional measures are not suitable for assessing the performance. Most of these specific metrics, such as precision and recall, exist for classification problems. The notions of precision and recall were adapted to regression problems with non-uniform relevance of the target values by Torgo and Ribeiro [19] and Ribeiro [18]. We will use the framework proposed by these authors to evaluate and compare our results. More details on this formulation can be obtained in [18].

Table 1: Regression algorithms, parameter variants, and respective R packages.

Learner	Parameter Variants	R package
MARS	$nk = \{10, 17\}, degree = \{1, 2\}, thresh = \{0.01, 0.001\}$	earth [20]
SVM	$cost = \{10, 150\}, gamma = \{0.01, 0.001\}$	e1071 [21]
Random Forest	$mtry = \{5, 7\}, ntree = \{500, 750, 1500\}$	randomForest [22]

All the described alternatives were evaluated according to the F-measure with $\beta = 1$, which means that the same importance was given to both precision and recall scores. The values of F_1 were estimated by means of 3 repetitions of a 10-fold Cross Validation process and the statistical significance of the observed paired differences was measured using the non-parametric pairwise Wilcoxon signed-rank test.

Optimization In the optimization sub-problem the objective was to assign to each state a certain amount of police officers, given the total budget, the total number of available officers, and the violent criminality predictions. The optimization was carried out in R software, with the package “lpSolve”.

The values for the population and the density are real values, obtained from the estimates for 2014 [23]. However, the total budget, the number of available police officers, and the individual cost of the officers by state were defined by us. Although they are not real values, they serve as proof of concept. The cost of each officer by state was chosen randomly, and uniformly, from the interval [5, 15] once, then the same values were used in all experiments. Additionally, the values for u_b and l_b were set to 0.12 and 0.08, while v_l and v_h were set to 0.125 and 0.7, respectively.

4.2 Results and Discussion

Prediction We started by examining the results obtained with all the parameters selected for the two pre-processed data sets, the three types of learners and the smoteR variants. All combinations of parameters were tested by means of 3 repetitions of a 10-fold cross validation process. Figure 1 shows these results.

We have also analysed the statistical significance of the differences observed in the results. Table 2 contains the several p-values obtained when comparing the SmoteR variants and the different learners, using the non-parametric pairwise Wilcoxon signed rank test with Bonferroni correction for multiple testing.

The p-value for the differences between the two data sets (with 30% and 50% of the features) was 0.17. Therefore we chose the data set with less features to continue to the optimization problem. This was mainly because of: i) the non statistical significant differences and ii) the smaller size of the data (less features can explain well the target variable, so we chose the most efficient alternative).

Table 2: Pairwise Wilcoxon signed rank test with Bonferroni correction for the SmoteR strategies (left) and the learning systems (right).

Strategies	none	S.o2.u0.5	S.o2.u1	S.o4.u0.5	Learners	svm	rf
S.o2.u0.5	1.3e-14	-	-	-	rf	<2e-16	-
S.o2.u1	< 2e-16	1	-	-	mars	0.077	<2e-16
S.o4.u0.5	2.3e-16	1	1	-			
S.o4.u1	< 2e-16	0.18	1	1			

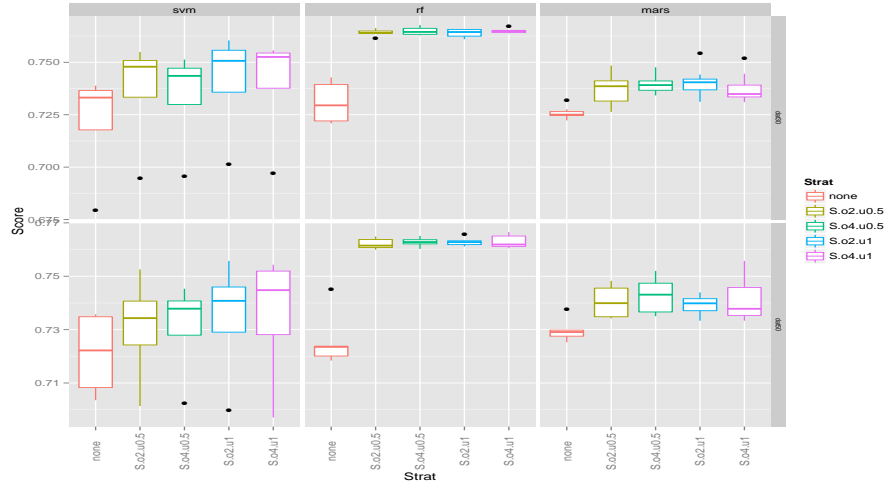


Fig. 1: Results from 3×10 -fold CV by learning system and SmoteR variant. (none-original data; S-smoteR; $ox-x \times 100\%$ over-sampling; $uy-y \times 100\%$ under-sampling)

Regarding the SmoteR strategy, Figure 1 and Table 2 provide clear evidence of the advantages of this procedure. Moreover, we also observed that the differences between the several variants of this procedure are not statistically significant. Therefore, we have opted for the variant which leads to a smaller data set and consequently a lower run time. For the optimization sub-problem we chose to use the smoteR variant with 200% of over-sampling percentage and 100% of under-sampling percentage. The learning system that provides a better performance is clearly the RF. With this learner, there is almost no differences among the several experimented variants.

Considering these results, we chose the following setting to generate a model for the optimization sub-problem:

- Pre-processing to remove missing values and select 30% of the most relevant features;
- Apply the smoteR strategy with parameters $k=5$; over-sampling percentage=200; under-sampling percentage=100;
- RF model with parameters: $mtry=7$; $ntree=750$.

After generating the model we obtained the predictions for the test set which was held apart to use in the optimization sub-problem. These predictions were used as input of the optimization task.

Optimization Several parameters were experimented. It was verified that with high budget and number of available officers, states with more criminality are assigned more officers. When the weight of the population increases, the most populated states, such as California, receive more police officers. When this weight

is decreased, those states lost officers, while, for instance, Vermont obtained the ideal number, although the population is one of the lowest

Table 3 shows the results of distributing 500,000 police officers, with a budget of 8,000,000, and $a = 1$. Figure 2 shows the same results in a map of the USA, where brighter red is associated with higher criminality, and the radius of the circles is proportional to the amount of officers assigned to the state. The color of the circle indicates which restriction limited the number of officers. Therefore, green means that the state received the ideal number, the minimum is represented in blue, yellow means that the budget of the state did not allow more officers, and white means that the state received a middle value of officers, which is less than the ideal or the maximum allowed by the budget, but higher than the minimum. It is possible to observe that ten states received the ideal number of officers. Some of them were associated with low or moderate levels of criminality, but the density or the population was high, such as New Jersey or Texas. Others are less populated, such as Oregon, but the ideal number of officers was also lower than other states constraint by the budget. The violent crime rate was particularly important in Kansas, since with a lower density and population, its budget allowed the state to receive the ideal number of officers. It is, also, possible to observe that the states with more violent criminality reached the number of officers allowed by their budget, such as Alabama or South Carolina. Accordingly, many states with less criminality received the minimum number of officers that they would allow (North Dakota), or values between the minimum and the ideal, without being constrained by the budget (Iowa). This behaviour may be desirable, since having too many officers in states with less criminality may be a waste of resources. The influence of the crime severity may be perceived when comparing Arizona with Nevada. The former has more population, higher density and budget than the latter, but received less officers because of the lower criminality rating.

According to the FBI [1], the region with more violent crime incidents is the South, followed by the West, Midwest and Northeast. It is interesting to notice that, in Figure 2, it was predicted more severe criminality for the southern states. These were the states that receive more police officers.

5 Conclusions

In this paper, we proposed a pipeline for predicting violent crime and a resources optimization scheme. Prediction encompasses feature selection through correlation and feature importance analysis, over-sampling of the rare extreme values of the target variable and regression. Among the evaluated learning systems, RF presented the best performance. This pipeline itself is one of the contributions of this work, given that, to the best of our knowledge, this problem in this data set was never approached as regression. Having the predictions, we propose a decision support scheme through the optimization of police officers across states, while taking into account the violent crime predictions, population, density and budget of the states. This contribution is presented as a proof of concept, since

Table 3: Distribution of 500,000 police officers by state, subjected to a total budget of 8,000,000.

State	Crime Prediction	Budget	Min. Off.	Ideal Off.	Dist. Off.	Cost
NJ	676.8	320390.4	1597	18614	18614	182659.9
PA	276.1	323802.5	1279	15991	1279	11413.7
OR	638.3	87076.7	678	7951	7951	59607.0
NY	893.1	504835.7	4445	49991	41007	504830.6
MO	381.3	12559.3	123	1503	123	1280.8
MA	408.1	233862.6	842	10283	842	5457.7
IN	726.6	164404.1	1247	14420	14420	162921.4
TX	821.1	648307.0	5643	64214	64214	477378.9
CA	850.2	948629.7	8369	94781	68083	948620.3
KY	655.6	104572.9	769	8996	8996	66277.2
AR	928.2	64349.2	691	7726	5885	64346.6
CT	355.4	146317.7	413	5090	413	6109.4
OH	542.8	90044.9	587	6997	587	8778.9
NH	312.6	33606.1	142	1760	142	2027.5
FL	1313.2	503583.7	6432	67716	52441	503581.4
WA	557.3	168051.3	1089	12954	12954	120927.3
LA	1721.0	109854.6	1994	19765	11893	109847.5
WY	528.1	1846.3	87	1036	87	778.2
NC	1315.6	247220.2	3221	33900	27118	247214.5
MS	1089.4	65686.5	808	8801	7149	65677.9
VA	851.7	208745.0	1798	20363	14106	208734.1
SC	1171.5	119505.3	1397	15028	10543	119498.9
WI	325.7	136544.5	629	7793	629	5271.4
TN	712.7	160819.6	1219	14128	10940	160817.3
UT	794.8	61723.4	599	6850	4253	61723.1
OK	488.6	86322.2	545	6560	545	5960.0
ND	342.0	6056.3	83	1026	83	618.2
AZ	500.5	155514.1	962	11554	962	11214.4
CO	791.7	121546.9	1087	12432	12432	84710.8
WV	551.2	39327.5	283	3371	283	3485.0
RI	440.9	110983.5	138	1680	138	1327.7
AL	1452.9	113590.5	1738	17915	7786	113576.2
GA	1254.6	248026.3	3120	33144	26054	248017.1
ID	444.8	28553.6	216	2616	216	2020.0
ME	275.9	23475.0	133	1663	133	1942.2
KS	1286.4	60751.9	920	9724	9724	52508.6
SD	568.5	8853.4	133	1585	1403	8851.3
NV	920.6	58244.7	656	7350	7350	56186.7
IA	556.7	67617.0	479	5696	2243	15039.1
MD	1271.9	191051.4	1872	19832	13217	191045.6
MN	862.1	125640.6	1191	13465	13465	96440.4
NM	835.6	39199.9	443	5031	3950	39199.7
DE	1161.9	48882.5	268	2891	2891	16497.7
VT	517.2	8881.3	92	1097	92	548.1
AK	932.3	64349.2	694	7752	7752	45227.2
DC	3044.8	926793.2	553	4612	4612	67407.4

some of the parameters were synthesized and may not correspond to the real scenario. Nevertheless, our results show an higher crime burden in states located in the southern part of the USA compared with the states in the north. For this reason, southern states tend to have an higher assignment of police officers. These predictions are in accordance with some national reports, and although some parameters of the optimization are not completely realistic, it seems to work as expected.

This work, although limited to the United States, can be easily applied to various other countries. So, as future work we consider that it would be interesting to apply the proposed framework in other countries or regions.

5. S. Shah, F. Bao, C.-T. Lu, and I.-R. Chen, "Crowdsafe: crowd sourcing of crime incidents and safe routing on mobile devices," in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 521–524, ACM, 2011.
6. 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*, pp. 231–238, Springer, 2012.
7. R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. S. Panahy, and N. Khanahmadliravi, "An experimental study of classification algorithms for crime prediction," *Indian Journal of Science and Technology*, vol. 6, no. 3, pp. 4219–4225, 2013.
8. S. Shojaee, A. Mustapha, F. Sidi, and M. A. Jabar, "A study on classification learning algorithms to predict crime status," *International Journal of Digital Content Technology and its Applications*, vol. 7, no. 9, pp. 361–369, 2013.
9. A. L. Buczak and C. M. Gifford, "Fuzzy association rule mining for community crime pattern discovery," in *ACM SIGKDD Workshop on Intelligence and Security Informatics*, p. 2, ACM, 2010.
10. M. A. Redmond and T. Highley, "Empirical analysis of case-editing approaches for numeric prediction," in *Innovations in Computing Sciences and Software Engineering*, pp. 79–84, Springer, 2010.
11. G. Donovan and D. Rideout, "An integer programming model to optimize resource allocation for wildfire containment," *Forest Science*, vol. 49, no. 2, pp. 331–335, 2003.
12. J. Caulkins, E. Hough, N. Mead, and H. Osman, "Optimizing investments in security countermeasures: a practical tool for fixed budgets," *Security & Privacy, IEEE*, vol. 5, no. 5, pp. 57–60, 2007.
13. P. S. Mitchell, "Optimal selection of police patrol beats," *The Journal of Criminal Law, Criminology, and Police Science*, pp. 577–584, 1972.
14. M. Daskin, "A maximum expected covering location model: formulation, properties and heuristic solution," *Transportation Science*, vol. 17, no. 1, pp. 48–70, 1983.
15. L. Li, Z. Jiang, N. Duan, W. Dong, K. Hu, and W. Sun, "Police patrol service optimization based on the spatial pattern of hotspots," in *Service Operations, Logistics, and Informatics, 2011 IEEE International Conference on*, pp. 45–50, IEEE, 2011.
16. L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, "Smote for regression," in *Progress in Artificial Intelligence*, pp. 378–389, Springer, 2013.
17. L. Torgo, P. Branco, R. P. Ribeiro, and B. Pfahringer, "Resampling strategies for regression," *Expert Systems*, 2014.
18. R. P. Ribeiro, *Utility-based Regression*. PhD thesis, Dep. Computer Science, Faculty of Sciences - University of Porto, 2011.
19. L. Torgo and R. P. Ribeiro, "Precision and recall in regression," in *DS'09: 12th Int. Conf. on Discovery Science*, pp. 332–346, Springer, 2009.
20. S. Milborrow, *earth: Multivariate Adaptive Regression Spline Models. Derived from mda:mars by Trevor Hastie and Rob Tibshirani.*, 2012.
21. E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, and A. Weingessel, *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*, 2011.
22. A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
23. U. S. C. Bureau, *Population Estimates [Online]*. Available: <http://www.census.gov/popest/data/index.html> [Accessed:23-01-2015], 2012.