Revista Brasileira de Ciências Policiais
ISSN: 2178-0013
ISSN: 2318-6917
steniosantos.sss@pf.gov.br
Academia Nacional de Polícia
Brasil

Using a Classification Model to Proper Deploy Police Patrol to face Bank Robbery in Northeast Brazil

Silva, Wellington Clay Porcino; Macedo, José Antônio Fernandes de; Neto, José Florêncio de Queiroz
Using a Classification Model to Proper Deploy Police Patrol to face Bank Robbery in Northeast Brazil
Revista Brasileira de Ciências Policiais, vol. 13, núm. 9, 2022
Academia Nacional de Polícia, Brasil
Disponible en: https://www.redalyc.org/articulo.oa?id=673470948012

Artigos

Using a Classification Model to Proper Deploy Police Patrol to face Bank Robbery in Northeast Brazil

Wellington Clay Porcino Silva
Polícia Federal, Brasil
Universidade Federal do Ceará, Brasil
wellington.wcps@pf.gov.br

José Antônio Fernandes de Macedo
Universidade Federal do Ceará, Brasil
jose.macedo@insightlab.ufc.br

José Florêncio de Queiroz Neto
Universidade Federal do Ceará, Brasil
florencio@insightlab.ufc.br

Redalyc: https://www.redalyc.org/articulo.oa?id=673470948012

Recepción: 22 Marzo 2021
Aprobación: 08 Agosto 2021

Abstract:

This article proposes a new approach to help police officers fight bank robberies, especially a violent type of crime named, in Brazil, "Novo Cangaço." Bank robbery is a massive problem in small towns all over Brazil and, particularly, in the Northeast. In this context, police managers face a complex challenge in deploying their patrols to cover huge areas. To cope with this problem, we propose a new approach using classification algorithms that use the probability of a bank robbery event based on territorial characteristics to deploy more police officers efficiently. We will also analyze geographical features to understand our model, explaining how they impact bank robberies events, using feature importance functions.

Keywords: police management, bank robbery, artificial intelligence models, geography, algorithms.

Resumo:

Este artigo propõe uma nova abordagem para auxiliar policiais no combate a assaltos a bancos, especialmente um tipo violento de crime denominado, no Brasil, "Novo Cangaço". O assalto a banco é um grande problema em pequenas cidades de todo o Brasil e, principalmente, no Nordeste. Nesse contexto, os gestores policiais enfrentam um desafio complexo ao desdobrar suas patrulhas para cobrir grandes áreas. Para lidar com esse problema, propomos uma nova abordagem usando algoritmos de classificação que usam a probabilidade de um evento de assalto a banco com base em características territoriais para implantar policiais com mais eficiência. Também analisaremos características geográficas para entender nosso modelo, explicando como elas impactam os eventos de assaltos a bancos, usando funções de importância de características.

Palavras-chave: gestão policial, roubo a bancos, modelos de inteligência artificial, geografia, algoritmos.

Resumen:

Este artículo propone un nuevo enfoque para ayudar a los policías a combatir los robos a bancos, especialmente un tipo de delito violento denominado, en Brasil, "Novo Cangaço". El robo de bancos es un problema masivo en los pequeños pueblos de todo Brasil y, en particular, en el Nordeste. En este contexto, los directores de policía enfrentan un desafío complejo al desplegar sus patrullas para cubrir áreas extensas. Para hacer frente a este problema, proponemos un nuevo enfoque utilizando algoritmos de clasificación que utilizan la probabilidad de un evento de robo a un banco en función de las características territoriales para desplegar más agentes de policía de manera eficiente. También analizaremos las características geográficas para comprender nuestro modelo, explicando cómo impactan los eventos de robos a bancos, utilizando funciones de importancia de características.

Palabras clave: gestión policial, robo de un banco, modelos de inteligencia artificial, geografía, algoritmos.

1. INTRODUCTION

One of Brazil's most significant public safety problems is how to face a critical kind of bank robbery known as "Novo Cangaço". It is an incredibly violent robbery, in which at least ten criminals take control of a small town, using extreme violence and hostages, to steal one or all banks in the town. Bank robbery occurs primarily, as said before, in small towns, with little or no police squads, which makes this type of crime challenging to be faced.

The name "Novo Cangaço" is derived from Cangaço, referring to the bands of armed criminals in northeastern Brazil, which appeared from the mid-nineteenth century, operating until the first decades of the twentieth century and whose greatest exponent is Virgulino Ferreira, known as Lampião. One of its characteristic forms of action was related to the looting of cities. It is exactly this way of acting that makes him similar to the phenomenon known as "Novo Cangaço" (MENESES, 2012)[SLdO1] . These criminal organizations have a great area as awareness space, representing huge areas containing many municipalities. It is very complex to deploy police patrols effectively because there are not enough patrols to cover all areas where they can act. Predictive Policing techniques, such as Risk Terrain Maps (RTM), found in the literature ((NATH, 2006), (CAVADAS; BRANCO; PEREIRA, 2015), (CAPLAN; KENNEDY, 2011), (CHAO HUANG et al, 2018), (ZHAO; TANG, 2017)) use regression models that more suitable to be used in micro areas located within a city instead of huge areas. Regression models usually establish a risk degree for each study area.

We apply a classification model to identify which municipalities have more probability of being attacked by a criminal organization specialized in bank robberies. Besides, use vector data for better visualization instead of raster data used when we employ regression techniques.

As a constraint, we will use the simplest model possible, with good results, because algorithms to use in Law Enforcement problems should be very easy to understand and to explain due the risk of bias and prejudices to specific groups. This explanation should be also the simplest possible, so Police Decision Makers can support them easily. Explainability is a huge problem specially for algorithms used in Law Enforcement because police decisions should be clearly understandable due its possible consequences, which can include even death of police officers or civilians.

We also established our main Domain Task (DT): to prioritize police resources deployment in large areas and tried to establish whether the model we are proposing is adequate to this DT, when we are facing criminals with a broader awareness space and specific targets as bank robbers, who usually commit their crimes in many municipalities.

To fulfill this goal, we asked specialists to select which one of the models presented, one using raster data, made with regression technique, and the other using our classification model, with vector data, is better to select where police decision-makers should deploy their resources.

2. PRELIMINARIES

During 2016 and the first months of 2017, in Rio Grande do Norte, a state on Brazil's northeast coast, faced a bank robbery spread with three different criminal organizations acting simultaneously. These criminals could attack almost every little town, as happens in "Novo Cangaço" crimes.

Police forces do not have many tools to face this challenge because they have few human and materials resources, making a hard decision to deploy specialized forces to face these criminal organizations, mainly because these gangs have a vast area of interest, covering almost all of the state.

To face the problem, police forces in that state created a Task Force, including Federal and State Civil Police Forces, which are investigative law enforcement agenciesand State Military Police, the ones who are responsible for deploying specialized patrols to give a fast response where and when a crime occurred. To combat the “Novo Cangaço” was difficult because they did not have many human resources, and the area was too big.

One possible solution was to use a Risk of Terrain Map, which are maps that use regression algorithms to calculate the risk that a micro place, like a small grid, has for an event to occur.

Nonetheless, this technique has been used with success in small areas as neighborhoods, as a result, is a fragmented map where each small grid has a different level of risk, when we deal with large areas, from a business point of view, it is hard to decide which grid prioritize when they are fragmented in small areas. In this kind of crime, as we focus in small towns, if we could predict which municipality had a greater risk, it would be easier to decide how to deploy patrols, because we already know, in each town, where we should go to give a fast response or even avoid the crime. Our goal in this paper is to make the police decision-making process better to face “Novo Cangaço” by developing a tool more adequate to evaluate the risk to this specific case.

3. PROBLEM STATEMENT

To Achieve this, we propose a problem stated as follows. Let M = {m1, m2, ..., m n} denote a set of municipalities from Rio Grande do Norte, where n is the number of municipalities and X = [X1, X2, ..., Xn] denotes the feature matrix of all municipalities, given a Municipality Mi and its features Xi, calculate the probability of a bank robbery event to occur in Mi.

4. RELATED WORK

To evaluate the risk of a crime occurring in a micro place and to deploy police patrols in areas of greater risk is an effective strategy of policing (BRAGA, PAPACHRISTOS, HUREAU, 2010). So, to achieve this goal, many models were created to predict future crime at micro places (CAPLAN; KENNEDY, 2011) (CHANEY; RATCLIFFE, 2008) (ZHAO; TANG, 2017).

Predicting crime is no longer just an academic question, but each day is more common in policing practices (PERRY et al, 2013). Crime Prediction algorithms can be classified according to the used method in three classes (WHEELER; STEENBEEK, 2020) (REINART; GREENHOUSE, 2018): a) Hot Spot maps; b) near repeat analysis; and c) regression-based methods.

Hot Spots maps are the most employed method by police forces worldwide and try to find places with a high concentration of crimes based on temporal and spatial patterns (CHANEY; RATCLIFFE, 2008).

Near repeat analysis is based on the observation that 31% to 76% will return to the same area if they succeed in committing a crime, which increases the risk of a crime happening in areas with a significant concentration of recent crimes (BERNASCO, 2008). To identify this recurrence, Johnson et al. (2007) suggest using Knox Test of Spatio-temporal clustering to identify places with high-risk rates due to repeated analysis. Another technique based on the near repeat analysis theoretical framework uses the Self-Exciting Points Process to crime prediction (MOHLER et al, 2011).

There are many Regression-based models employed in crime forecasting as K-mean clustering combined with a weighting algorithm in a geographical approach (NATH, 2006), Random Forest Regression (CAVADAS; BRANCO; PEREIRA, 2015), Risk Terrain Modeling (CAPLAN; KENNEDY, 2011) and even Deep Neural networks-based algorithm, as DeepCrime (CHAO HUANG et al, 2018) and temporal-spatial correlations in crime-based model (ZHAO; TANG, 2017).

All the methods stated above have one thing in common: they try to predict risk in micro places, small areas, and not in municipalities or bigger spaces, as a whole state. As a consequence of this characteristic, they are not suitable for the business problem we want to address here: to face bank robbery in the whole state of Rio Grande do Norte. To support police decision-makers in this task, we propose a different approach to use a classification model to calculate the probability of such a crime in a municipality based on geographical features. The resulting map would be cleaner by using discrete values instead of continuous representation, and it will help police managers deploy more efficiently their resources in such scenarios.

5. FEATURE ENGINEERING

Spatial features are indeed related to crimes (CHANEY; RATCLIFFE, 2008). Bank Robberies and especially "Novo Cangaço" cases are obviously impacted by geographical characteristics of towns targeted by criminal organizations.

According to Costa et al. (2016), distance from great cities, a dense road network and proximity to states borders are spatial factors which contribute to a town to be targeted by this kind of criminal organizations. Silva (2019) also states that distance from previous bank robberies; population under 50.000 inhabitants; distance from states borders and density of highways and cell phone networks have an impact in the criminal decision process in choosing the most suitable target for the attack. Still according to Costa et al. (2016) and Silva (2019), all features described above enhance the possibilities of a municipality being targeted by bank robbers.

We will use these territorial characteristics as features Xi to model the probability of a municipality Mi to be attacked by this type of criminal organizations, using a classification algorithm.

We have had to prepare our dataset from files described in table 1:

TABLE 1 –
Files used to prepare the dataset

Description	Source
Bank Robberies in 4 Brazilians States from 2014 to 2016	Federal Police
RN's Highways	Federal Police
Brazilian States	IBGE
Brazilian Municipalities	IBGE
Brazilian Municipalities Downtown	IBGE
Brazilian Population in 2019 (Estimation)	IBGE
Cell Phones Companies services by Municipality	ANATEL

So, each feature was calculated as shown below:

a) min_dist_to_borders: Minimum Euclidean distance from municipality downtown to state borderline.

b) min_dist_to_highway: Minimum Euclidean distance from municipality downtown to the nearest highway.

c) Roubos_Sim/Nao: This is our target label, where 1 means that a bank robbery has happened in that municipality and 0 that it has not occurred. We used this variable as a target so we can establish which municipality has a greater risk of being attacked by this kind of criminal organization.

e) dist_roubos: Euclidean distance from municipality downtown to nearest bank robbery.

f) Rede_3G_4G: Feature where 1 means that municipality has 3G or 4G cell phone service, else receives a 0 as value.

g) Pop_classes: Municipalities were classified in four groups: 1 from 0 to 2,100 inhabitants; 2 from 2,100 to 89,709 inhabitants; 3 from 89,710 to 386,342 inhabitants and 4 more than 386343 inhabitants. We decide to use this classification because as pointed in Silva (2019), population impacts in bank robbery probability depends more on how a municipality is classified as small, medium or big than on the exact number of inhabitants. To classify, we used Head/Tail Breaks Algorithm (JIANG, 2012) because we have a heavy-tailed distribution, with a great number of municipalities with a small population.

Once we have prepared the dataset, from exploratory data analysis we conclude:

FIGURE 1:
Correlation Matrix

	min_dist_to_borders	min_dist_to_highway	Nr_Roubos	Roubos_Sim/Não	dist_roubos	Pop_classes
min_dist_to_borders	1	-0,12	-0,083	-0,025	-0,34	-0,053
min_dist_to_highway	-0,12	1	0,035	-0,063	0,024	-0,0099
Nr_Roubos	-0,083	0,035	1	0,47	-0,18	0,74
Roubos_Sim/Não	-0,025	-0,063	0,47	1	-0,39	0,22
dist_roubos	-0,34	0,024	-0,18	-0,39	1	-0,14
Pop_classes	-0,053	0,0099	0,74	0,22	-0,14	1

Correlation Matrix

a) Number of crimes and population classes has a very high correlation as we can see in figure 1, so we decide to drop the number of crimes.

b) We have a very imbalanced dataset. We have 148 municipalities in class 0 (no robbery) and only 19 in class 1 (at least one robbery), so just 11,38% of our observations are in class 1. To address this question, we employed an oversampling technique named Synthetic Minority Over-Sampling Technique (SMOTE) (CHAWLA et al, 2002) with a number of neighbors equal to 8. Before the oversampling operation, we randomly selected 15% of our dataset as a test dataset, which will be used to evaluate our model afterwards.

c) We dropped Rede_3G_4G because all municipalities have 3G or 4G cell phone networks.

Our final dataset was composed of 256 rows, equally divided in class 1(at least one robbery) and 0 (no robbery), and 5 columns: min_dist_to_borders, min_dist_to_highway, dist_roubos, Pop_classes, Roubos_Sim/Nao.

It is important to addresses that, even with an oversampled dataset, we still have a small number of events, due that Bank Robberies are not as usual as others crimes, such as thefts or even homicides, in Brazil.

6. MODELING

Our final dataset was composed of 256 rows, equally divided in class 1 and 0, and 5 columns: min_dist_to_borders, min_dist_to_highway, dist_roubos, Pop_classes, Roubos_Sim/Nao.

We normalized our data using the z score method to avoid possible outliers' issues. Then, we applied a yeo-johnson power transformation to make our feature values more normal-distribution-like because our first approach would use logistic regression, which is a more straightforward method achieving good results being explainable using the feature importance method.

To train the model, we performed ten-fold cross-validation using each of the algorithms stated in Table 2.

Table 2: Comparing Models

We also established we would use as target metric Recall, because we would like to emphasize true positive and because false positive would not have great impacts.

As we can see in table 2, Logistic Regression, CatBoost, K Neighbors Classifiers, Naïve Bayes, Linear Discriminant Analysis and Ridge Classifier has scored 1.0 in Recall, our target metric. When we have algorithms with same performance, we should choose the simplest, with the lowest computational cost. In our use case, other very important factor to analyze, before choose an algorithm is explainability. To support Law Enforcement operations, it is crucial for a machine learning model to be explainable to police decision makers and even to judges and attorneys.

So, using these information we selected a Logistic Regression algorithm due it is the simplest,the most explainable and it has achieved 1.0 score in Recall, the best score obtained in our target metric.We did not choose CatBoost Classifier due it is a tree-based model which is not easy to explain. Algorithms to use in Law Enforcement problems should be very easy to understand and to explain due the risk of bias and prejudices to specific groups. Even though we can use shapley value to learn feature importance in tree models, we decided to keep it as simple as possible, so it would be easy to be supported by police decision makers.

Then we have tuned our logistic regression model using a random search and as target metric Recall with Pycaret library, and obtained as parameters C equals to 3.931, penalty L2 and solver lbfgs number of iterations maximum equals to 100.

After tuning our model, we have kept our mean Recall score equals to 1.0 and improved our mean AUC from 0.925 to 0.950.

We also can see our results using the ROC Curve in figure 2.

Figure 2 – ROC Curve

To demonstrate how well our model can classify our data we have plotted figure 3, decision boundary. As we can see our model is able to classify our data with a good performance.

Figure 3: Decision Boundary

7. EVALUATING

To determine whether the proposed model generalizes well, and not do a good work in our training dataset, it is important to randomly divide our dataset in two: a training dataset to be used in training our model and a test dataset, that we keep until the very end and this last one is used to evaluate our model. We can choose different percentages to split our dataset in training and test, because this decision is all about a trade-off between estimation of generalization error and withholding valuable information to train our model (RASCHKA; MIRJALILI, 2017). In our case, we split the dataset in 15%-85% to test and training dataset.

Then, nonetheless results we obtained in the previous phase; it is important to see how well our model can classify unseen data. To do that, we randomly selected 15% of our original dataset before over-sampling, as stated above, so this data could be used as a test dataset.

So, we passed this unseen data in our model and our target metric, i.e., Recall Score has kept itself equals to 1.0, which demonstrates how well our model can generalize.

8. EXPLAINING THE MODEL

As stated before, explainability is a desirable characteristic in a machine learning model. This issue is even more important when algorithms are used to support police decisions, due its possible consequences.

So, to understand how are model works, it is important to demonstrate how each feature impacts our result because it makes it easier for police decision makers to trust our model.

Then, we applied the Feature Importance function to demonstrate how each one of the independent variables contributes to the result and plotted it in figure 4.

Figure 4 – Feature Importance Plot

As we can see, distance from previous bank robberies is the most important feature. It was expected because, according to environmental criminology, a criminal tends to act in an awareness space where he feels comfortable (CHANEY; RATCLIFFE, 2008). So, it is expected that we can identify clusters of crimes in such conditions, even when we are working on a greater scale and not in micro places, as we usually employ this type of technique.

Population is the second most important feature, which is also expected because population has a great correlation with the number of police officers in town, as pointed out by Silva (2019).

9. VISUALIZATION OF THE RESULTS

Once we have evaluated our model, we should visualize results. As we were working with spatial data, it was fundamental that we represented using a map, which is mandatory when we deal with spatial data.

In figure 5, we can see a map, using vector data instead of raster data. In each municipality, our model, using Logistic Regression, stablished a probability of such place to suffer a bank robbery.

As stated before, we decide to propose this visualization instead of a traditional one, based in a regression model in which each pixel has a value, as we can see in figure 6.

We adopted this visualization because we considered that using vector data associated with each city would be a better approach to our Domain task, which is to prioritize police resources deployment. Our hypothesis is that it is easier to decide where to deploy our resources to face bank robberies criminal organizations in municipalities basis than analyzing small parts of terrain, as we can see in figure 6, below. And to create this map, we should use a classification algorithm, instead a regression one, more often used in predictive policing problems, and more suitable to small areas.

We have decided to show our results using a continuous scale as legend, where the value is equal to the probability of such Municipality of Rio Grande do Norte suffers an attack of “Novo Cangaço”.

Figure 5 – Bank Robberies Probabilities in Rio Grande do Norte

As we already said, we used vector data to make the map, showing bank robbery probability in each municipality.

Figure 6 - Risk Terrain Map for Bank Robbery in Brazil’s Northeast Region (SILVA,2019)

The map shown in figure 6 was made using statistical inference with raster data that presented the risk of a bank robbery.

To establish if the approach using vector data (figure 5) is better than the traditional one, using raster data (figure 6) for larger areas, we have sent a form to a hundred sworn police officers with experience in fighting bank robberies.

In this form we have shown figures 6 and 7 asked them which visualization they think is better to prioritize police resources deployment in large areas when they are facing criminals with a broader awareness space.

These specialists were divided into 60 investigators, from Federal and Civil police forces and 40 from Military Police, specialists in patrolling. All of them work nowadays or have worked a long time in units that are specialized in fighting bank robberies.

From the 100 specialists who have received the form, 75 have answered. Analyzing the results, we can see that 73 (97.3%) have preferred the map using vector data, as shown in figure 5, and 2 (2.7%) police officers preferred the visualization in figure 6. With such a result, we can state that for our DT, a classification map (figure 5) is better than a regression map (figure 6).

So, using our approach, police decision makers can choose more efficiently deploy their patrols and investigative resources to face these criminal organizations, because they now can prioritize municipalities with higher risk. This, allied to intelligence information, can be a great weapon to help police officers in fighting against "Novo Cangaço".

In figure 7, we have an aggregated view of the answers of specialists, from google forms, who demonstrates how robust our visualization proposal is to deploy police resources to face criminals who act in larger areas, like bank robbers.

Figure 7 – Specialist's Answers

10. CONCLUSIONS

In this paper, we presented a new approach to deal with a well-known problem: crime prediction. Previous techniques proposed to solve this problem are based on regressions approaches and in micro-place crime concentration.

Nevertheless, if we face a problem that touches all over a state? How to try to predict in an actionable way where to deploy police patrols when you have a whole state to look for? That the challenge police decision-makers must face when dealing with "Novo Cangaço" criminal organizations. Moreover, it is for them we have decided to create this model.

Nonetheless, we will continue this research by employing this model in real-world police operations, which includes deploying it as a service in a geo-intelligence application used by police officers in Brazil and researching other features that help improve our model.

ACKNOWLEDGES

We thank the Polícia Federal and the Computer Science Department of Federal University of Ceará (DC/UFC) for supporting this work. This paper was developed during Author 1 post-doctoral internship at DC/UFC.

AUTHOR´S BIOGRAFY

Wellington Clay Porcino Silva - Delegado de Polícia Federal. Pós-Doutor em Ciência da Computação pela UFC. Doutor em Geografia pela UFRN. Mestre em Sistemas de Informações Geográficas pela Universidade Nova de Lisboa. Mestre em Operações Militares pela EsAO/EB. Especialista em Ciência Policial e Inteligência pela ANP. Especialista em Atualização Pedagógica pela UFRJ. Especialista em Direito Penal e Processual Penal pela Universidade Gama Filho. Foi chefe do SIP/SR/DPF/RJ, da Divisão de Doutrina de Inteligência Policial da DIP e da Divisão de Repressão aos Crimes Ambientais da DICOR, Superintendente da Polícia Federal em Roraima e no RN, Diretor de Operações e de Gestão e Integração da informação na SENASP/MJ. Atualmente exerce o Cargo de Coordenador de Gestão Estratégica/PF. Possui interesse em avaliar artigos sobre gestão de polícias, indicadores, formulação de políticas públicas de segurança e emprego de de tecnologia na segurança pública e na gestão policial

José Antonio Fernandes de Macedo - É professor adjunto do departamento de computação da Universidade Federal do Ceará. Concluiu mestrado e doutorado pela Pontificia Universidade Católica do

Rio de Janeiro nos anos 2001 e 2005, respectivamente. Durante o doutorado passou 8 meses pesquisando na TELECOM Bretagne - École Nationale Supérieure des Télécommunications Bretagne na Franca e em seguida 10 meses na École Polytechnique Fédéral de Lausanne (EPFL). Realizou seu pos-doutorado na EPFL no periodo de 2006-2009, onde coordenou as atividades de pesquisa junto ao projeto europeu GeoPKDD (www.geopkdd.eu). O foco de sua pesquisa é processamento de dados em larga escala em nuvens computacionais. Bolsista de Produtividade em Pesquisa do CNPq - Nível 2. Possui ainda um especial interesse no desenvolvimento de algoritmos de mineração de dados e aprendizagem de máquinas para processamento de grandes volumes de dados (Big Data).

José Florêncio de Queiroz Neto - Doutor em Ciência de Computação pela Universidade Federal do Ceará. Pesquisador do Insightlab do Departamento de Computação da Universidade Federal do Ceará.

REFERENCES

BERNASCO, W. Them Again?: Same-Offender Involvement in Repeat and Near Repeat Burglaries. ***European Journal of Criminology*****, 5, p. 411-431, 2008.**

BRAGA, AA; PAPACHRISTOS, AV; HUREAU, DM. The concentration and stability of gun violence at micro places in Boston, 1980-2008. ***Journal of Quantitative Criminology*** **26, p. 33-53, 2010.**

CAPLAN, J M; KENNEDY, L W. Risk terrain modelling: Brokering criminological theory and GIS methods for crime forecasting. ***Justice Quarterly*** **28, p. 360-381, 2011.**

CAVADAS, B. BRANCO, P; PEREIRA, S. Crime Prediction Using Regression and Resources Optimization. DOI: 10.1007/978-3-319-23485-4_51, 2015.

CHANEY, S; RATCLIFFE, J. ***Gis and crime mapping*****. West Sussex: John Wiley & Sons, 2008.**

CHAO HUANG ***et al. DeepCrime:*** **Attentive Hierarchical Recurrent Networks forCrime Prediction. CIKM'18. Turim: Association for Computing Machinery, 2018.**

CHAWLA, N. V. ***et al*****. "Smote: Synthetic Minority Over-sampling Technique.** ***Journal of Artificial Intelligence Research.*** **16, p. 321-357, 2002.**

COSTA, C. A. V. ***et al. Novo Cangaço:*** **sua incidência no Estado do Pará e sua relação com o tráfico de drogas. Em Segurança e Defesa: Conflitos, Criminalidade, Tecnologia da Informação, por S. S. Almeida, E. M. L. S. Ramos e C. A. N. Chagas, 197-206. Belém: Edições UFPA, 2016.**

JIANG, B. Head/tail Breaks: A New Classification Scheme for Data with a Heavy-tailed Distribution. ***The Professional Geographer.*** **p.1-13, 2012.**

JOHNSON, S. D. ***et al*****. Prospective crime mapping inoperational context: Final report. Home** ***Office Online Report*** **19/07. London, 19 de Jul., 2007.**

MOHLER, G ***et al*****. Self-Exciting Point Process Modeling of Crime.** ***Journal of the American Statistical Association.*** **106, p.100-108, 2011.**

NATH, S. V. Crime pattern detection using data mining. ***Web Intelligence and Intelligent Agent Technology Workshops, IEEE/WIC/ACM International Conference*****. IEEE, p. 41-44, 2006.**

PERRY, W. L. ***et al. Predictive policing:*** **the role of crime forescasting in law enforcements operations. Washington: RAND Corporation, 2013.**

RASCHKA, Sebastian; MIRJALILI, Vahid. Python Machine Learning. 2 ed. Birmingham: Packt, 2017.

REINHART, A; GREENHOUSE, J. Self-exciting point processes with spatial covariates: Modelling the dynamics of crime. ***Journalof the Royal Statistical Society*****: Series C (Applied Statistics) 67, p.1305-1329, 2018.**

SILVA, W. C. P. S. ***Espaço geográfico e criminologia*****: topologia de segurança versus topologia do crime - uma análise da gestão de segurança do território e roubo a bancos no Nordeste. Tese apresentada ao Programa de Pós-Graduação em Geografia do Departamento de Geografia da Universidade Federal do Rio Grande do Norte para a obtenção do título de Doutor em Geografia. Natal, 2019.**

WHEELER, A P; STEENBEEK, W. Mapping the risk terrain for crime using machine learning. *J Quant Criminol* (10.31235/osf.io/xc538), 2020. https://doi.org/10.1007/s10940-020-09457-7.

ZHAO, X; TANG, J. *Modeling Temporal-Spatial Correlations for Crime Prediction.* CIKM'17. Singapore, p.497-506, 2017.