International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 10 Issue 3, March, 2021, Impact Factor: 7.751

Using the Machine learning approach to predict theft and stealing

Chaitanya Sharma

Student, Delhi Public School, Gurgaon, Haryana, India

ABSTRACT

Robbery and thieving are two wrongdoings against the land that has an enormous societal impact. Their expectation brings down the rate of exploitation and a sense of vulnerability in society. The purpose of this study is to compile a record that allows for the prediction of rehash offences by lawbreakers in these kinds of wrongdoings, to help patterns of precaution practises. A collection of machine learning data from the Crime Analysis and Investigative Focus System (CIFS) of the Regional Public Prosecutor's Office in Biobío, Chile, was prepared to obtain the case. The data provided by robberies and thefts submitted somewhere in the range of 2012 and 2017 in the city of Concepción. The outcomes show a portrayal of recurrent wrongdoers in these kinds of wrongdoing and a repeated file that takes into consideration a more prominent decisiveness in the forecast of recidivism than the technique that is presently being utilized.

Keywords: Machine Learning, AI, Crime Prediction, Knowledge Discovery of Databases (KDD)

1. INTRODUCTION

The Chilean Public Prosecutor's Office is a self-ruling yet progressive entity whose role is specifically to facilitate the evaluation of constituent unlawful conduct and, in the way required by law, to carry forth unauthorized public activities. There is a Criminal Analysis and Investigative Focus System (CIFS) to enhance the design of administration and the impact of criminal proceedings, whose task is to increase the criminal arraignment of property violations and different misdoings with strong social significance. The CAIFS of the Provincial Public Prosecutor's Office in Biobío develops the positioning of the hoodlums with the most important risk of recurrence within that year in[1-3] order to monitor the analytical work zeroed in on reducing the quantity of high social significance wrongdoings, for example, violations against the land, at the beginning of every year. This positioning helps the knowledgeable assets to concentrate on the lawbreakers with a more influential risk of recurrence.

The number of criminal incidents that a person has had over the most recent year is used to build up this positioning. The higher the number of felony cases, the greater the probability of recidivism by the lawbreaker. While the number of criminal cases is vital to recidivism, it does not depend entirely on this aspect, but rather on a separate arrangement of elements linked to the personality's individual and socio-demographic characteristics and the values associated with the person's criminal background. The Chilean Public Prosecutor's Office is now recording statistics on individuals who have committed wrongdoing and is mainly identified with the relevant parts of their chronicled particular criminal offences. The goal of this analysis is to draw up a list that quantifies the degree of recidivism of an individual, taking into account their criminal background so that the Regional Public Prosecutor's Office in Biobío can essentially zero in criminal prosecution work and improve the adequacy of the expected assets for this work.

Given the uncertainty of the recurrence wonder and the challenge of constructing a particular example that characterises and forecasts it, the recurrence record can be obtained with the use of knowledge mining models. Information mining can eliminate designs present in the information that are not clear[4] to be used later for forecasting purposes; for example, the director of a person.

2. AI IN CRIME PREDICTION

The advanced society creates a quickly developing amount of information, producing new issues and conceivable outcomes [5]. This has implied a significant test for law application, where data assumes a pertinent part for investigators, who must do a detailed and productive examination of violations [6–7]. When considering the routine activities of a lawbreaker, vast amounts of information are produced, which makes it hard to recognize wrongdoing

through regular information examination [8]. Because of the previously mentioned, the studio of misconduct ends up being one of the most relevant fields for the use of information mining. The information that it could deliver to examine, control and forestall wrongdoing makes it a valuable instrument to help police work [9–10].

Substance extraction, Cluster analysis, rules of the association, designation and analysis The amount of criminal prosecutions that an individual has taken in the most recent year is used to set up this placement. The higher the number of criminal cases, the greater the risk of recidivism by the lawbreaker. While the number of criminal cases is essential to recurrence, it does not focus solely on this factor, but rather on the separate structure of the elements related to the individual and socio-demographic features of the victim and the beliefs correlated with the criminal history of the offender. The Chilean Public Prosecutor's Office is also recording information on people who have committed wrongdoing and is primarily identified with the related aspects of their chronic criminal offences. The purpose of this study is to draw up a list that quantifies the degree of recidivism of a person, taking into account their criminal history, so that the Regional Public Prosecutor's Office in Biobío can be effectively null in criminal cases. Media networks [11-12].

In this analysis, the classification approach would be used to generate a list of recidivism. There is a broad variety of models within this method that are otherwise called learning machines[13]. Machines learning learn the amount of the overall enclosed architecture in the details and then use it to make another prediction. The expectation is to do away with a vault or a view of a class or class previously characterised, e.g. Repeat or Not Repeat. The forecasts that machine learning conveys are an opportunity somewhere in the region between zero and one, defined as certainty. This value will be used as a recurrence pointer, where a price close to one will be associated with a high likelihood of recurrence and a value close to zero will be identified with a low chance of recurrence.

Decision Tree: refers to an organisation table of a progressive tree layout. It helps to discover installations in high-dimensional spaces and in issues that mix all and mathematical knowledge. In this model, each hub is a function on which the test is conducted. The branches coming from the corner speak to the groups of the property that display the test result. Each terminal hub or leaf indicates at the stage to which a record or impression is allocated.

Naive Bayes: Based on the estimation of the probability of finding a position with a class, using Bayes Theorem for the assessment of dependent probabilities. This dictates the restricting possibility that a record will have a level area, considering all of the variables that portray it. This AI assumes autonomy between the elements of the agreement.

It appears to be found in the published audit that Decision Tree and Naive Bayes are commonly used in settings identified with the commission of various kinds of misconduct. Therefore, they were selected to be included in this test.

3. MATERIAL AND METHODS

The Knowledge Discovery of Databases (KDD)[28] was used for the recognisable evidence of the organisation of the critical variables and the planning of the templates for the acquisition of the recidivism record. This technique is based on a series of stages, the main purpose of which is the extraction of enclosed knowledge inside data sets[29]. This method is represented as a non-unimportant disclosure period of information and potentially valuable details inside the information stored in the data collection. It's anything but a scripted process; it's an iterative loop that deeply explores the tremendous amount of knowledge required to determine the designs.

This strategy is one of five stages. The key step is an educated decision, where the sources of evidence and the kind of data to be used are resolved. The necessary information for the review shall be excluded from the information source (s).To select the data, the variables available in the creation period must be fully understood and the variable to be predicted must be settled. The next step consists in the planning of the knowledge base, to provide more accurate results, which contributes to a stronger opportunity for forecasting. This process consolidates the analysis of lost information, contradictory information, and the investigation of out-of-range information.

The third stage includes the adjustment and determination of the causes. The growth of factors requires every loop that adjusts the state of the plot. Factors are modified to deliver new elements that enhance the data that will be used to prepare the model so that it has stronger prescient power. After the transition, the loop proceeds to discern those who better forecast the element of suspense by means of methods that evaluate each factor's constructive staff[30]. With this, it's possible to create less abstract models that will render things easy to clarify.

The fourth stage is information mining. Methods that remove the essential example from information are used at this point; the characterization process is used explicitly in this study. The fifth stage in the loop is a prescient evaluation of the presentation. This involves the calculation of the AI being used. For this purpose, the results obtained from the application of the test set in the prepared model shall be used. The consequences are summarised by a network called Matrix and Ambiguity. For example, if two classes are thought of, type 1 that detects a persistent perp and level 0 that recognises a non-recurring suspect, the Ambiguity Matrix[31] will have the form seen in Table 1.

Table 1: Confusion Matrix

Classes	Real value		
		1	0
Forecasted Value	1	TP	FN
	0	FN	TP

In the Confusion Matrix seen in Table 1, TP refers to the class 1 components that were successfully predicted by the model or real positive rate, and FN refers to the stage 1 elements that were incorrectly predicted by the model or bogus positive figure. TN refers to the class 0 components that were successfully predicted by the model, or the real negative rate, and FP speaks to the category 0 components that were falsely assumed by the model or false positive rate.

The Uncertainty Matrix is given with the following presentation measures[32].
Accuracy: speaks to the full scope of the predictions that have been correctly characterised.

$$Accuracy = (TP+TN)/(TP+FP+FN+TN)$$

THE REVIEW IS THE LEVEL OF PERCEPTIONS THAT HAVE A PLACE WITH CLASS 1 AND WERE ACCURATELY ARRANGED BY THE MODEL.

$$Recall = True\ Positive/\ (True\ Positive/False\ Negative)$$

PRECISION: IS THE LEVEL OF ACCURATELY GROUPED COMPONENTS AS CLASS 1 OF THE MEMBERS DELEGATED CLASS 1.

$$Precision = True\ Positive/\ (True\ Positive/False\ Positive)$$

EXACTNESS GAUGES THE OVERALL APPEARANCE OF THE MODEL, WHILE RECALL AND ACCURACY ACHIEVE A MORE CAREFUL APPROXIMATION OF HOW THE MODEL DIRECTLY FORECASTS THE CLASS OF EXCITEMENT ALLUDED TO ABOVE, A PREDICTION THAT THE CLASSIFIER CONVEYS AN OPPORTUNITY IN THE RANGE OF ZERO AND ONE. THIS VALUE CAN BE USED AS A RECURRENCE RECORD, WHERE A VALUE CLOSE TO ONE IS ASSOCIATED WITH A HIGH RISK OF RECURRENCE AND A COST NEAR ZERO TO A LOW PROBABILITY OF RECURRENCE. THE LIST, DEMANDED IN A PLUMMETING STRUCTURE, PROVIDES A RECIDIVISM POSITIONING THAT ENCOURAGES CRIMINAL INVESTIGATION WORK TO ZERO ON THOSE INDIVIDUALS WITH THE HIGHEST SCORES.

4. KDD Process Application to Obtain the Recidivism Index

A. Dataset Selection

The details provided by the CAIFS of the Public Prosecutor's Office of Biobío relates to 12,222 robberies and burglaries reported sometime between 2012 and 2017. Each record is represented by seven ascribes, as seen in Table 2.

Table 2: Attributes Of The Original Database

Attribute	Description
RUT (national identity N°.)	Corresponds to the unique identifier for each offender that has committed a crime.
Crime	Corresponds to the name of the crime committed by the offender.
RUC (criminal case N°.)	The unique criminal case code assigned to the offender.
Date of crime	The date on which the crime was committed.
Criminal convictions	Indicate if the person has had any convictions.
Gender	Sex of the accused.
Date of birth	Represents the date of birth of the offender.

The unique identifier of record in Table 2 is drawn from the RUT, RUC, and Date of Crime ascribes.

B. Pre-preparing of Data

At this point, each property was inspected in order to detect atypical and missing details. For each function, the identifiable proof of atypical knowledge was rendered using the three-sigma rule[33]. Out-of-range material was considered, others were forgotten, and some were substituted. The replacement of atypical and incomplete data was done using the Hot Deck process, which substitutes the missing character estimate with that of a property from a comparative record[34]. By evaluating the date of birth quality, a distinction was drawn between persons under the age

of 14 who do not fall under the jurisdiction of the Public Prosecutor's Office. Records of individuals younger than 14 years of age have been disposed of.

C. Trait Transformation

The RUT is a specific identification for any convicted party for which a recidivism file is requested. Since the information base does not display a single RUT for each record, it is important to adjust the data collection to the correct configuration. This reform allowed 11 new ascribes to be made out of the first five (RUT, Crime, RUC, Date of Crime, and Date of Birth). Credits were rendered using the Recency Frequency Monetary (RFM) process, which is commonly used in the Client Division[35]. In this approach, Recency refers to the period that has elapsed since the client's last transaction, Frequency speaks to their recurrence of sales, and Monetary says to the accumulated amount of the properties. As far as this review is concerned, the notion of the Recent is conveyed in terms of the slipping of time after the last person committed misconduct. The concept of frequency is spoken to by the usual.

Table 3: Description Of The Created Attributes

	Attribute	Description
A	RUT (national identity N°.)	Corresponds to the unique identifier for each offender that has committed a crime.
B	Age at last crime	Corresponds to the age that an offender had when they committed the last crime recorded in the database.
C	Age during last year	Corresponds to the age during the last year of records from the database.
D	Days since last crime	The number of days that have elapsed since the last crime committed by an offender.
E	Quantity of crimes	Represents the number of crimes that an individual has committed.
F	Quantity of crimes in the last period	Counts all of the crimes that an offender has had during the last year in the database.
G	Quantity of RUC (criminal cases)	Represents the number of criminal cases in which an offender has participated.
H	Average number of accomplices per RUC	Represents the average number of individuals with whom an offender participates in criminal cases.
I	Crime accomplices	Is the number of different individuals with whom an offender is linked to in the different criminal cases in which they have participated.
J	Average number of days between crimes	Corresponds to the average difference in days between one crime and another. The investigation only considers repeat offenders.
K	Average crimes per year	Corresponds to the average number of crimes committed per year.
L	Quantity of RUC in the last period	Number of criminal cases recorded by an individual during the last period.
M	Gender	Sex of the offender.
N	Conviction record	Indicates if the person has had a conviction.
O	Recidivism	Indicates if an offender has committed a crime against property in the year following the last year of study.

The number of days between the offence and the total number of wrongdoings per year. The numerical concept is the gross amount of transgressions of the offending party, the number of offences in the most recent year, and the number of associated court proceedings. Additional intrigue credits that are inconsistent with this approach have also been made. The totality of the ascribes is seen in Table 3.

The relation mechanism was established after the properties were rendered in Table 4. The relation distinguishes whether there is a direct dependence between the mathematical ascribes that have been made. A high association between the two credits infers that these two explains a miracle that, compared, like this for the planning of the model, each of those features would be left with a correlation record more remarkable than or equal to ±0.9. The ascribes who were killed were as follows:

- Age during a year ago (C)
- Quantity of violations in the last time frame (F)
- Average wrongdoings every year (K)

D. Characteristic Selection

After the shift in the knowledge base, the more important credits for the recidivism prediction are picked. Its prescient intensity of recurrence characterises the importance of each quality. For the planning and testing of templates, ascribes with the highest constructive power are considered. The Chi-squared calculation was used to calculate the visual strength of each property. This calculation is a proportion of the dependency between some property and the trait to be expected. The more remarkable the estimate of the scale, the greater the dependency between these properties; the more

important the valuation of the property would be over the predicted quality. This technique is only suitable for straight credits, so for its execution, it was important to figure out each of the mathematical acronyms, selecting the number of classes that enhanced its dependency on the trait to be predicted. This classification and the visionary strength of each character.

Table 4: Correlation Matrix Between Created Attributes

Attributes	B	C	D	E	F	G	H	I	J	K	L
B	1	0.994	−0.233	0.101	0.136	0.226	−0.154	−0.109	0.185	0.101	0.188
C	0.994	1	−0.129	0.078	0.072	0.196	−0.148	−0.114	0.154	0.078	0.128
D	−0.233	−0.129	1	−0.235	−0.628	−0.325	0.063	−0.042	−0.298	−0.235	−0.613
E	0.101	0.078	−0.235	1	0.403	0.856	0.009	0.166	−0.074	1	0.434
F	0.136	0.072	−0.628	0.403	1	0.409	−0.010	0.077	−0.004	0.403	0.926
G	0.226	0.196	−0.325	0.856	0.409	1	−0.127	0.103	0.081	0.856	0.523
H	−0.154	−0.148	0.063	0.009	−0.010	−0.127	1	0.825	−0.101	0.009	−0.06
I	−0.109	−0.114	−0.042	0.166	0.077	0.103	0.825	1	−0.042	0.166	0.06
J	0.185	0.154	−0.298	−0.074	−0.004	0.081	−0.101	−0.042	1	−0.074	0.063
K	0.101	0.078	−0.235	1	0.403	0.856	0.009	0.166	−0.074	1	0.434
L	0.188	0.128	−0.613	0.434	0.926	0.523	−0.06	0.06	0.063	0.434	1

Table 5: Categories By Attribute And Predictive Power According To The Chi-Squared Statistic

Attribute	Categorization	Statistical value Chi-squared
Quantity of RUC (criminal cases)	5 categories: [1,2] – [3,4] – [5,6] – [7,8] – >=9	143.709
Quantity of RUC in the last period	4 categories: Does not have-[1,2]-[3,4]->4	96.073
Days since last crime	5 categories: [1, 219] – [220,438] – [439,657] – [658,876] – >877	82.734
Average number of days between crimes	8 categories: [1,20] – [21,40] – [41,60] – [61,90] – [91,150] – [151,300] – [301,530] – >=531	82.006
Age at last crime	6 categories: [14,24]-[25,34]-[35,44]-[45,54]-[55,64]->64	47.383
Quantity of crimes	4 categories: [1,5] – [6,10] – [11,15] – >=16	44.451
Conviction record	2 categories: YES – NO	6.807
Average number of accomplices per RUC	5 categories: [0,1] – [1,2] – [2,3] – [3,4] – >4	3.426
Average number of days between crimes	5 categories: [0,1] – [1,2] – [2,3] – [3,4] – >4	1.520
Gender	2 categories: Feminine, Masculine	0.414

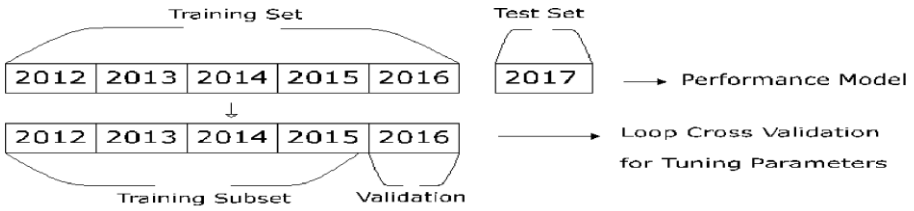

Fig. 1: Hold-out cross-validation technique for time series.

E. Information Mining

Three machines of learning referenced in the previous section were prepared and evaluated using the formulas in the Scikit-Learn Python programming language library[36]: Decision Tree Classifier, Naive Bayes, and Multilayer Perceptron. In their planning, the different boundaries were acclimatised to enhance their prescient exhibition. In order to prepare them, it was necessary to change the details, so that the number of records in both classes was equal. Via this adjustment, the models have fairly gained popularity with the overall example that represents the variable to be expected, towards a greater component class. The knowledge equilibrium was rendered using the Synthetic Minority Oversampling Process, which produces new documents for the minority class within its neighbourhood[37]. For the planning period, the acceptance and acquisition of time limits from the hold-out cross-approval process are used [38–39]. This technique considers the fleeting distribution of knowledge in the planning and testing set, as seen in figure 1.

Figure 1 indicates that the planning collection formed by the documents from 2012 to 2016 is split into the preparation (2012–2015) and acceptance (2016) subsets for the changing of boundaries.

This boundary adjustment is done using an iterative cycle until the optimal boundary for the model is defined. With the boundaries modified, the paradigm is being prepared with a series of preparations, and recidivism or non-recidivism is expected in 2017. Using this expectation, the prescient presentation of each model is acquired. The effects of the anticipation of models are investigated in the following section.

5. INVESTIGATION OF RESULTS

Table 6: Predictive Performance Of The Machines Learning Used

	Classification model		
Performance measure	Decision tree classifier	Naive Bayes	Multilayer perceptron
Accuracy	59%	76%	71%
Recall	87%	66%	73%
Precision	27%	37%	32%

After checking various quality blends, using the constructive powers from Table 5 as a basis of an insight, the best exhibition in each preparation model was obtained by thinking about the five ascribes with the best creative power: the quantity of RUC (Criminal Cases),

Sum of RUC in the last time span, Days after prior misconduct, Total number of days between breaches, and eventually commit to mischief. The strongest constructive power that each model has accomplished is nitty-gritty in Table 6.

Table 6 appears to demonstrate that Naive Bayes obtains the highest overall efficiency at an accuracy rate of 76 per cent, followed by the neural organisation Multilayer Perceptron with 71 per cent, and finally the Decision Tree Classifier at a precision rate of 59 per cent. Despite these overall results, we can see that, in the detailed prediction of the human recidivism class, the models are more dominant in Recollection than in Accuracy. This differentiation reveals that, for the most part, models better anticipate individuals who are actually recurrent wrongdoers and who are not.

From the organisational point of view of the forensic prosecutors at CAIFS from the Biobío Public Prosecutor's Office, there is a more significant error in the expense of predicting whether the wrongdoer will re-outrage (type I blunder) than is anticipated if the accused party does not re-affront (type II mistake). The cost of class 1 blunders is psychological, as precaution calculations would not be taken for an individual than would be re-insulted. The effect of Form II errors would be the time and assets allocated to safeguard steps for an individual who is not re-insulting. The self-assurance of recidivism placement would be viewed as a characterising metric to define the best model. For this reason, the graphic called the Lift Map is used, which indicates the true positive rating of the variety (precisely clustered chronic wrongdoers) as shown by the rate of the people within the positioning. Figure 2 displays the Benefits Map for each of the models.

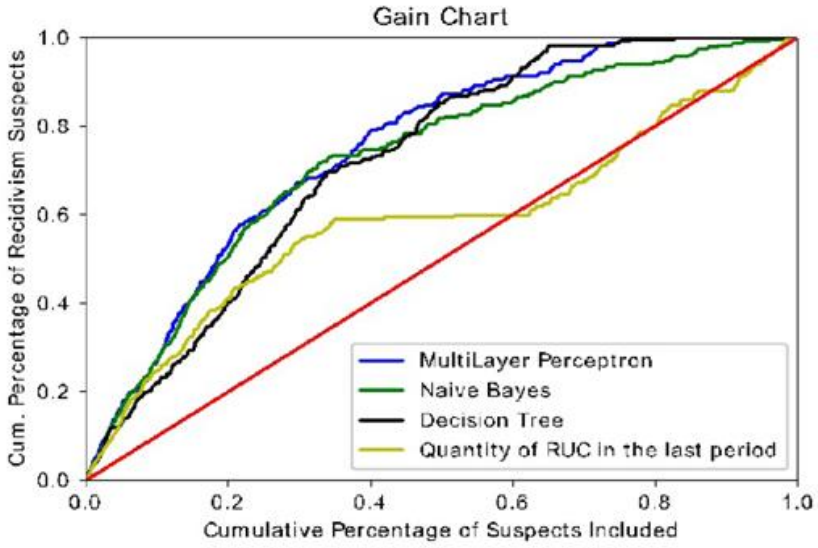

Fig. 2: Machines Learning Performance and the current method used by the Public Prosecutor's Office.

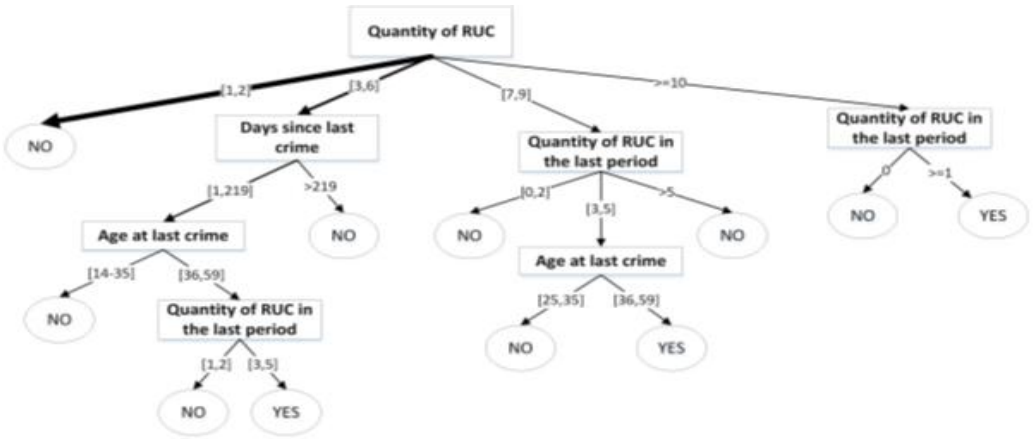

Fig. 3: Decision tree created to characterize the recidivism in thefts and burglaries.

In the diagram seen in Figure 2, the upper presentation infers a bend that is closer straight ahead (0.1), which is connected to the most important number of individuals arranged at the highest seating positions. It can be shown very well that the best exhibition is carried out by the Multi-Layer Perceptron, lastly preceded by Naive Bayes Decision Tree Classifier. The totality of the models reveals an unrivalled display of the methodology used by the CAIFS of the Public Prosecutor's Office of Biobío, which consists of arranging wrongdoers who focus on the number of criminal cases in the previous year. This method fits wonderfully on individuals with a large number of felony incidents, where the curve is similar to the source. This can be seen in more significant detail in Figure 3, which displays the Decision Tree generated by planning the effects of the Decision Tree Classifier AlgorithmThe option tree in Figure 3 shows that, as a rule, it is practicable to represent any persistent defendant by using four credits: cumulative number of incidents, number of patients in the last time period, days after prior misconduct, and eventually agree to mischief when the accused parties report at least two illegal bodies of evidence relating to infringements of property that they do not re-outrage. At the point where the number of cases reported by the wrongdoer is high, with at least one case recorded in the last time frame, the person will be re-annoyed in the next time frame. These two branches approve the measures currently being used by the CAIFS of the Public Prosecutor's Office of Biobío, which can be essentially characterised as the more significant (less) the number of cases recorded, the more prominent (less) the likelihood of recidivism, particularly in the case of a small number of issues.

The situation turned out to be more unpredictable with the middle amount of wrongdoings that occur in the focal sections of the tree. These branches laid out the accompanying basic guidelines for recidivism:

- If the wrongdoer is in a range between three and six criminal cases, fewer than 220 days have elapsed since the last wrongdoing, his/her age is somewhere between the area between 36 and 59, and the criminal cases in the range of three and five have been carried out in the most recent year, the person would be re-irritated.
- Where there are seven and nine court convictions reported by a wrongdoer, some of which have been in the area of three and five in the most recent year. In addition, three of them are in the vicinity of 36 and 56; they would be re-irritated in the following time frame. Even though the options tree has general rules that lead an individual to have a high probability of becoming a chronic criminal, the neural organisation must be used as a model for generating a list of recidivism that prefers to transfer to the wrongdoer.

CONCLUSION

The recidivism file proposed in this analysis relies on a significant number of general details given by the CAIFS of the Public Prosecutor's Office of Biobío concerning breaches. In the light of these data, a model that offers an indication of persistent guilty parties may be prepared and used as a recidivism measure beginning from one time to the next. Naive Bayes is the greatest AI by and wide execution. In any case, the importance of this file is the recidivism placement, and the neural organisation Multi-Layer Perceptron is the positioning of the model that thinks the persistent guilty parties higher in the rankings.

REFERENCES

[1] De la Fuente H, Mejías C. Análisis econométrico de los determinantes de la criminalidad en Chile. Política Criminal. 2011; 6(11), 192–208. http://dx.doi.org/10.4067/S0718-33992011000100007
[2] Archwamety T, Katsiyannis A. Factors related to recidivism among delinquent females at a state correctional facility. Journal of Child and Family Studies. 1998; 7(1), 59–67. https://doi. org/10.1023/A:1022960013342

[3] Katsiyannis A, Archwamety T. Factors related to recidivism among delinquent youths in a state correctional facility. Journal of Child and Family Studies. 1997; 6(1), 43–55. https://doi. org/10.1023/A:1025068623167
[4] Han J, Kamber K, Pei J. Data mining: concepts and techniques. 3rd edn. The Morgan Kaufmann. 2011. https://www.academia.edu/download/43034828/Data_Mining_Concepts_ And_Techniques_3rd_Edition.pdf.
[5] Cocx T, Kosters W. A distance measure for determining similarity between criminal investigations. In: Industrial conference on data mining. Springer, Berlin, Heidelberg. 2006; 511–525. https://doi.org/10.1007/11790853_40
[6] Chen H, Chung W, Xu J, Wang G, Qin Y, Chau M. Crime data mining: a general framework and some examples. Computer. 2004; 37(4), 50–56. https://doi.org/10.1109/MC.2004.1297301
[7] De BruinJ, Cocx T, Kosters W, Laros J, Kok J. Data mining approaches to criminal career analysis. In: Sixth international conference on data mining (ICDM'06). 2006; 171–177. https:// doi.org/10.1109/ICDM.2006.47
[8] Thongtae P, Srisuk S. An analysis of data mining applications in crime domain. In: 2008 IEEE 8th international conference on computer and information technology workshops. 2008; 122– 126. https://doi.org/10.1109/CIT.2008.Workshops.80
[9] Keyvanpour M, Javideh M, Ebrahimi M. Detecting and investigating crime by means of data mining: a general crime matching framework. Procedia Computer Science. 2011; 3, 872–880. https://doi.org/10.1016/j.procs.2010.12.143
[10] Nath S. Crime pattern detection using data mining. In: 2006 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology workshops. 2006; 41–44. https://doi.org/10.1109/WI-IATW.2006.55
[11] Hassani H, Huang X, Silva E, Ghodsi M. A review of data mining applications in crime. statistical analysis and data mining: The ASA Data Science Journal. 2016; 9(3), 139–154. https:// doi.org/10.1002/sam.11312
[12] Troncoso F, WeberR. A novel approach to detect associations in criminal networks. Decision Support Systems. 2020; 128, 113–159. https://doi.org/10.1016/j.dss.2019.113159
[13] Kotsiantis S, Zaharakis I, Pintelas P. Machine learning: a review of classification and combining techniques. Artificial Intelligence Review. 2006; 26(3), 159–190. https://doi.org/10.1007/ s10462-007-9052-3
[14] Appavu S, Pandian M, Rajaram R. Association rule mining for suspicious email detection: a data mining approach. In: 2007 IEEE intelligence and security informatics. 2007; 316–323. https://doi.org/10.1109/ISI.2007.379491
[15] Kirkos E, Spathis C, Manolopoulos Y. Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications. 2007; 32(4), 995–1003. https://doi. org/10.1016/j.eswa.2006.02.016
[16] Yu C, Ward M, Morabito M, Ding W. Crime forecasting using data mining techniques. In: 2011 IEEE 11th international conference on data mining workshops. 2011; 779–786. https:// doi.org/10.1109/ICDMW.2011.56
[17] Bhowmik R. Detecting auto insurance fraud by data mining techniques. Journal of Emerging Trends in Computing and Information Sciences. 2011; 2(4), 156–162. http://citeseerx.ist.psu.edu/ viewdoc/download?doi=10.1.1.302.3602&rep=rep1&type=pdf
[18] Fuller C, Biros D, Delen D. An investigation of data and text mining methods for real world deception detection. Expert Systems with Applications. 2011. 38(7), 8392–8398. https://doi. org/10.1016/j.eswa.2011.01.032