Crime Analysis and Prediction using Machine Learning

Conference Paper · September 2020

DOI: 10.23919/MIPRO48935.2020.9245120

1 author:

Olta Llaha
South East European University

Crime Analysis and Prediction using Machine Learning

Olta Llaha

South East European University/Faculty of Contemporary Sciences and Technologies, Tetovo, North Macedonia
E-mail: ol29064@seeu.edu.mk

Abstract **- Data mining and machine learning have become a vital part of crime detection and prevention. The purpose of this paper is to evaluate data mining methods and their performances that can be used for analyzing the collected data about the past crimes. I identified the most appropriate data mining methods to analyze the collected data from sources specialized in crime prevention by comparing them theoretically and practically. Some attributes of this dataset are, gender, age, employment status, crime place. Methods are applied on these data to determine their effectiveness in analyzing and preventing crime. Evaluations on the data showed that the method with a higher performance is "Decision Tree". This was achieved by some performance measures, such as the number of instances correctly classified, accuracy or precision and recall, that has brought better results compared to other methods. I come to the conclusion that the data mining methods contribute to the predictions on the possibility of occurrence of the crime and as a result in its prevention.**

Keywords - Machine Learning, Prediction, Crime Analysis, Data Mining

I. INTRODUCTION

The increase in crime data recording coupled with data analytics resulted in the growth of research approaches aimed at extracting knowledge from crime records to better understand criminal behavior and ultimately prevent future crimes.

Crime is a complex social phenomenon that has grown due to major changes in society. Law enforcement agencies need to learn the factors that lead to an increase in crime tendency. To curb this, there is always a need for strategies and policies to prevent crime. As a result of technology development, science and information, data mining and artificial intelligence tools are increasingly prevalent in the law enforcement community.

Law enforcement agencies face a large volume of data that needs to be processed and turned into useful information, and data mining can improve crime analysis by helping to predict and prevent it. By processing criminal data, law enforcement agencies can use models that may be important in the crime prevention process.

The use of data mining accelerates data analysis, and analysts can examine existing data to identify patterns and trends of crime. This paper is structured as follows: Section. 2 describes the relationship that exists between data mining, machine learning and criminology. The methodology and description of the dataset are described in Section. 3. Sections. 4 and 5, represent a theoretical description of the methods and algorithms that will be applied practically to our data. Section 6 presents the results of the application of algorithms and an explanation for the algorithm with the best results. In sect. 7 the conclusions and future work are discussed.

II. USING DATA MINING AND MACHINE LEARNING IN CRIMINOLOGY

Criminology is an area where the scientific study of crime and criminal behavior focuses. This is one of the most important areas when applying data mining techniques that can produce significant results [1].

Crime analysis, as part of criminology, is tasked with exploring and discovering crime and its relationship with criminals. Law enforcement is a process that aims to identify the characteristics of crime. Identifying crime characteristics is the first step in developing further analysis. The high volume of crime data and the complexity of the relationships between them have made criminology an appropriate field for applying data mining techniques [2].

Data mining can be used to examine many large datasets involving a large set of variables beyond what a single analyst, or even an analytical team or task force, can consider correct, whereas machine learning uses neural networks, predictive model and automated algorithms to make the decisions. Like any other problem solving method, the task of data mining begins with a problem definition. The identification of the data mining problem enables the determination of the data mining process and the modeling technique. Machine learning is a subfield of data science that deals with algorithms able to learn from data and make accurate predictions [3]. Data mining gives law enforcement agencies the opportunity to learn about crime trends, how and why crimes are committed. Using data mining methods and machine learning improves crime analysis and help reduce and prevent crime.

III. DATA AND METHODOLOGY

I compare theoretically and practically data mining methods to discover the most appropriate method for our data. The methods were compared by applying machine learning algorithms to concrete data in the WEKA "Waikato Environment for Knowledge Analysis" [4] environment. The implemented algorithms are: Simple Logistic, Logistic, Multilayer Perceptron, Naive Bayes, Bayes Net, SMO, C4.5.In data collection step I am collecting data from law enforcement agencies. The collected data is stored into database for further process. They relate to the areas where crime and perpetrator data occur.

The dataset is made up of 100 records or instances.

The name of the dataset	Number of examples	Number of input attributes	Number of possible classes	Total number of attributes	Values that are missing
Crime Data	100	6	2	7	0

Table 1. Dataset details

The variables or attributes of this dataset are: age (from 17 to 55 years old), gender, education (middle school. high school, university) employment status (whether employed or not), civil status (whether married, single, or divorced), the area where the crime occurred (urban or rural) and whether the person who committed the crime was previously convicted or not. Crime dataset is in CSV format.

IV. CLASSIFICATION METHODS

Classification is a data mining technique that categorizes data in order to assist in more accurate predictions and analyzes [5, 6]. It is one of the data mining methods that aims to analyze very large datasets. It is used to derive patterns that accurately define the important data classes within the data set. Classification consists in predicting a given result based on a given input [6]. Classification algorithms attempt to detect relationships between attributes that would make it possible to predict the result. They analyze the input and produce a prediction.

A. *Artificial Neural Networks*

Neural networks are an area of Artificial Intelligence (AI) based on the inspiration from the human brain. I use them to find data structures and algorithms for learning and classifying data. By applying neural network techniques, a program can learn from the examples and create an internal set of rules for classifying different inputs. Artificial Neural Networks (ANNs) are capable of predicting new observations from existing observations. A neural network consists of interconnected processing elements also called units, nodes, or neurons [5].

All processes of a neural network are performed by this group of neurons or units. Each neuron is a separate communication device, making its operation relatively simple. The function of one unit is simply to receive data from other units, as a function of the inputs it receives to calculate an output value, which it sends to other units. In artificial neural networks, neurons are organized in layers which process information using dynamic state responses to external inputs [6]. The Multilayer Perceptron (MLP) is a feed-forward artificial neural network model that maps sets of input data to a set of appropriate outputs [7]. In a feed-forward neural network, the input signal traverses the neural network in a forward direction from the input layer to the output layer through the hidden layers.

B. *Naive Bayes Classifier*

Bayesian classification represents a supervised learning method as well as a statistical classification method. It assumes a high-probability underlying model, which allows us to determine in principle the uncertainties for the model, thus determining the probability of the results. The Naive Bayes Classifier technique is based on the Bayesian theorem and is used especially when the dimensionality of the inputs is high [5, 8]. Naive Bayes classifier is a term in Bayesian statistics dealing with a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. Bayesian classification provides practical learning algorithms and prior knowledge, here the observed data can be combined. Bayesian classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates the apparent hypothetical probability. The algorithm works as follows. Bayes' theorem offers a way to calculate the probability of a hypothesis based on our prior knowledge.

Likelihood Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability Predictor Prior Probability

$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- $P(c|x)$ is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
- $P(c)$ is the prior probability of *class*.
- $(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

Class (c) is independent of the values of other predictors. Naïve Bayes Classifier can be trained effectively in supervised learning [8]. After calculating the conditional probability for a different number of hypotheses, I can solve the hypothesis (class) with the highest probability. An advantage of the Naive Bayes classifier is that it requires a small amount of training data to calculate the parameters (mean and variance of the variables) needed for the classification [8]. Because the independent variables are assumed, then only the discrepancies of the variables for each class need to be determined and not the full matrix distribution. The Naive Bayesian classifier is fast and incremental can deal with discrete and continuous attributes, has excellent performance and can explain its decisions.

C. Support Vector Machine

Support Vector Machines are based on the concept of decision making plans that set the boundaries of decisions. A decision plan is one that divides a group of objects that have different class memberships. Classification tasks that are based on the dividing lines between different class membership objects are known as hyper-plane Classifiers [9]. SVMs are a set of related supervised learning methods used for classification and regression. Support Vector Machine (SVM) is primarily a classification method that performs classification tasks by constructing hyper-plane in a multidimensional space. The SVM uses statistical learning theory to search for a regularized hypothesis that fits the available data well without over-fitting. SVM also supports regression and classification techniques and can handle multiple continuous and categorical variables [9].

The efficiency of SVM-based classification is not directly dependent on the dimension of the classified entities. SVM can also be extended to learn nonlinear decision functions by first projecting the input data into a high dimensional space using kernel functions and formulating a linear classification problem in that space. SMO (Sequential Minimal Optimization) implements John C. Platt's sequential minimal optimization algorithm for training a Support Vector classifier using polynomial or RBF(Radial Basis Function) kernels [9].This implementation globally replaces all lost values and transforms nominal attributes into binary ones. It can be seen that the choice of kernel function and best value of parameters for particular kernel is critical for a given amount of data. It also normalizes all attributes by default.

D. The decision tree

The decision tree is a method in which data is presented in a tree structure based on the values of their attributes. It splits the data in the database into subsets based on the values of one or more fields. This process will be repeated for each subgroup recursively until all instances are a node in a single class. The result of the decision tree is a tree-shaped structure that describes a series of decisions given at each step [5, 6]. These decisions are then considered as rules for the classification task. The algorithms commonly used to construct decision trees are; ID3 and C4.5.

The ID3 (Iterative Dichotomiser 3) algorithm [10] induces classification models, or decision trees, from data. It is a supervised learning algorithm that is trained by examples for different classes. After being trained, the algorithm should be able to predict the class of a new item. ID3 identifies attributes that differentiate one class from another. All attributes must be known in advance, and must also be either continuous or selected from a set of known values. For instance, temperature (continuous), and country of citizenship (set of known values) are valid attributes. To determine which attributes are the most important, ID3 uses the statistical property of entropy [10].

The C4.5 algorithm [11] overcomes this problem by using another statistical property known as information gain. Information gain measures how well a given attribute separates the training sets into the output classes. This algorithm has input in the form of training samples and samples. Training samples in the form of sample data that will be used to build a tree that has been substantiated. C4.5 algorithms are algorithms result of the development of the algorithm ID3 [11]. C4.5 algorithm works by grouping several training sample data that will result in a decision tree based on the facts on the training data.

V. ASSOCIATION RULES AND REGRESSION

Association Rule is one of the most important canonical tasks in data mining and probably one of the most studied techniques for pattern discovery. Association rules are if/then statements that help to uncover relationships between unrelated data in a database, relational database or other information repository [12]. Association rules are used to find the relationships between the objects which are frequently used together [12]. Association Rules identify the arguments found together with a given, event or record: "the presence of one set of arguments brings the presence of another set". This is how rules of type are identified: "if argument A is part of an event, then for a certain probability argument B is also part of the event" [13]. The objective of the association rule was to discover interesting association or correlation relationships among a large set of data items. Support and confidence are the most known measures for the evaluation of association rule interestingness.

While classification provides categorical, discrete labels, regression has continuous function values. So regression is used mainly to predict missing numeric data values rather than discrete class labels. Regression analysis is a statistical methodology often used for numerical prediction, although there are other methods for doing this [14]. Regression also involves identifying the distribution of trends based on available data. For this purpose regression trees can be used as well as decision trees whose nodes have numerical values instead of categorical values. Linear regression is a mathematical technique that can be used to make a numerical data set by creating a mathematical equation [14]. On the other hand logical regression estimates the probability of verifying an event under certain circumstances, using the factors observed together with the occurrence of the event [14].

VI. EXPERIMENTAL RESULTS

To conduct this study I used WEKA [4] software based on the approach and familiarity with its use. WEKA is a collection of machine learning algorithms for data mining tasks. It contains tools for data pre-processing, classification, regression, association rules, and visualization. It can be used to detect the various hidden patterns in our dataset and find the most determining data factors.

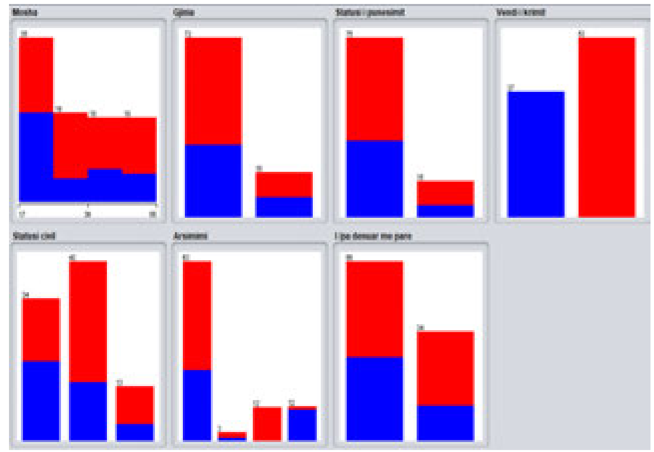

Figure. 1. Pre-processed data visualization

Experiments are done by using cross-validation on default option folds = 10. Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. The process is repeated 10 times for each fold. Performance indicators are given on the following Table 2.

Method	Algorithm	Correctly Classified Instances	Incorrectly Classified Instances	Recall	Precision	F-Measure
Regression	Simple Logistic	68 (68%)	32 (32 %)	0.680	0.680	0.665
	Logistic	71 (71%)	29 (29 %)	0.710	0.707	0.707
Bayes Classifier	Naive Bayes	73 (73%)	27 (27%)	0.730	0.737	0.732
	Bayes Net	72 (72%)	28 (28%)	0.720	0.725	0.721
SVM	SMO	67(67%)	33 (33 %)	0.670	0.666	0.666
Decision tree	C4.5	76 (76%)	24 (24 %)	0.760	0.762	0.761
Artificial Neural Network	Multilayer Perceptron	63 (63%)	37 (37%)	0.630	0.637	0.632

Table 2: Comparison of the results of the algorithms applied in WEKA

In this paper I used some algorithms (Table 2) and among them is C4.5 algorithm, which is a Decision Tree algorithm. This algorithm is clear and easy when I used it to interpret the results. The model construction is done by modifying the parameter values and this algorithm classifies crime data with a higher accuracy than other algorithms of data mining methods. I converted our data to format. The C4.5 algorithm was implemented in this data.

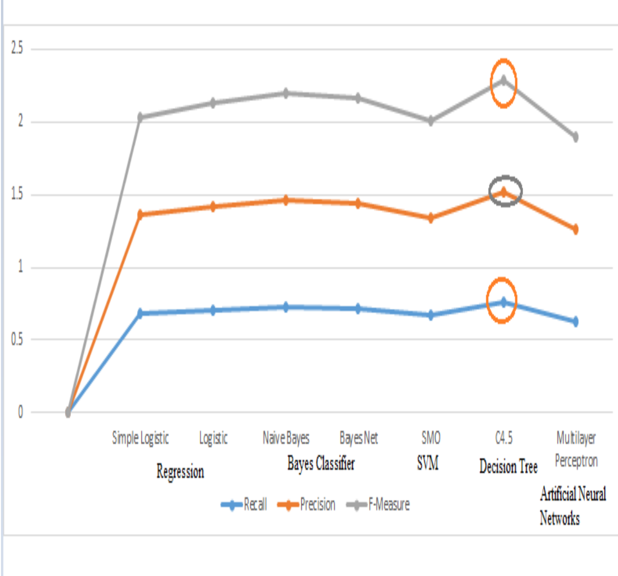

Figure 2: Performance of algorithms

The C4. 5 algorithm for building decision trees is implemented in WEKA as a classifier called J48. J48 has the full name weka.classifiers.trees.J48. What came out of this algorithm: the visualization and the decision tree are presented in Figure 3 and Figure 4.

```
=== Summary ===

Correctly Classified Instances          76               76      %
Incorrectly Classified Instances        24               24      %
Kappa statistic                          0.5106
Mean absolute error                      0.3194
Root mean squared error                  0.3997
Relative absolute error                 65.5331 %
Root relative squared error             80.973  %
Total Number of Instances              100

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.738    0.224    0.705      0.738    0.721      0.511    0.829     0.740     urbane
                 0.776    0.262    0.804      0.776    0.789      0.511    0.829     0.842     rurale
Weighted Avg.    0.760    0.246    0.762      0.760    0.761      0.511    0.829     0.799

=== Confusion Matrix ===

  a  b   <-- classified as
 31 11 |  a = urbane
 13 45 |  b = rurale
```

Figure 3: C4.5 (J48) Classifier

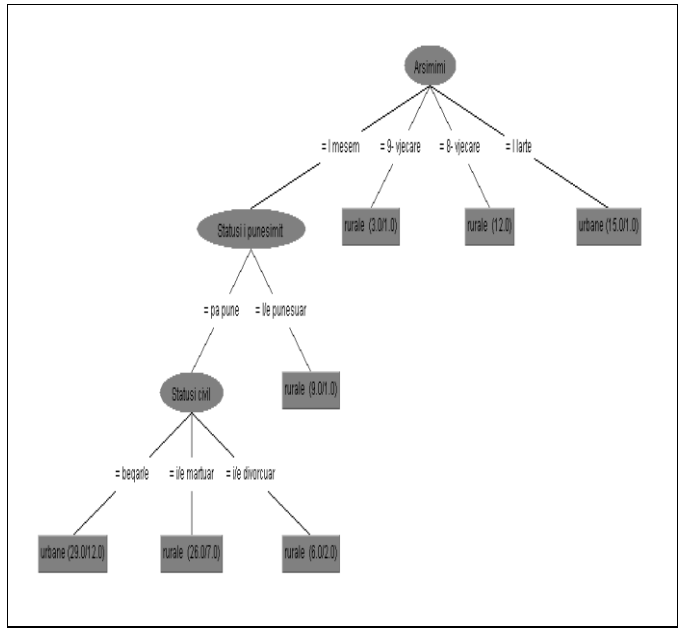

Figure 4: Decision Tree

Figure 3 shows the result of implementing the C4.5 algorithm. It shows that the number of correctly classified instances is 76 with a percentage of 76% and the number of incorrectly classified instances is 24, so 24%.

F-measure is a measure of a test's accuracy. It considers both the precision and the recall of the test to compute the score: precision is the number of correct positive results divided by the number of all positive results returned by the classifier, and recall is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive).

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

The results of this algorithm for recall and precision values are respectively 0.760 (recall) and 0.762 (precision).

$$\text{F-Measure} = \frac{2 * Recall * Precision}{Recall + Precision}$$

- True positive (TP): correct positive prediction
- False positive (FP): incorrect positive prediction
- True negative (TN): correct negative prediction
- False negative (FN): incorrect negative prediction

F-measure after the application of the algorithm has the value 0.761.

The implementation of this algorithm has classified the crime data based on the dataset attributes as e.g. the place where the crime occurred (urban areas, rural areas) where: the number of correctly, classified instances, the accuracy or precision and recall have the highest values compared to other algorithms of data mining methods.

Figure 4 shows the visualization of the decision tree which is generated by the implementation of the C4.5 algorithm. Through the decision tree generated I understand in which areas more crimes occur, as well as the characteristics of the people who committed the crimes. Having this information helps law enforcement agencies to create policies or make decisions about areas where the crime rate is higher.

VII. CONCLUSION AND FUTURE WORK

The purpose of this study is to examine crime analysis through the applicability of data mining methods in the process of crime prediction and prevention. The results of experiments conducted in this research by implementing algorithms of data mining methods have revealed that these methods are applicable in the process of crime prediction. The decision tree as a data mining classification method has classified crime data at an accuracy rate of 76%. This method has shown promising results for the problem of crime prediction as the accuracy rate is high in the experiments performed. Furthermore, the decision tree seems more viable due to the fact that in contrast to other algorithms, it expresses the rules explicitly. These rules can be expressed in human language so that anyone can understand them. The use of machine learning and data mining in crime analysis is important because data mining methods can be used in the decision making process. Decision making is very important in crime prevention in order to decide accurate actions and law enforcement strategies. Through our data analysis law enforcement agencies can create strategies, operating in areas where most crimes occur. In the future extension of this study some models will be created for predicting the crime hot-spots that would help the deployment of police to places of crimes. Algorithms' behavior changes will be looked at when more data is added. I also plan to look into developing social link networks of criminals, suspects and gangs. I also intend to implement this study to an integrated enterprise software that will be created.

REFERENCES

[1] K. Zakir Hussain, M. Durairaj and G. R. J. Farzana, "Criminal behavior analysis by using data mining techniques," IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), Nagapattinam, Tamil Nadu, 2012, pp. 656-658.

[2] Keyvanpour, Mohammad & Javideh, Mostafa & Ebrahimi, Mohammadreza. (2011). Detecting and investigating crime by means of data mining: A general crime matching framework. Procedia CS. 3. 872-880. 10.1016/j.procs.2010.12.143.

[3] Ioannis Kavakiotis OlgaTsave Athanasios Salifoglou Nicos Maglaveras Ioannis Vlahavas Ioanna Chouvarda, Machine Learning and Data Mining Methods in Diabetes Research, Computational and Structural Biotechnology Journal Volume 15, 2017, Pages 104-116

[4] Frank, Eibe & Hall, Mark & Holmes, Geoffrey & Kirkby, Richard & Pfahringer, Bernhard & Witten, Ian & Trigg, Len. (2010). Weka-A Machine Learning Workbench for Data Mining. 10.1007/978-0-387-09823-4_66.

[5] Pang-Ning Tan; Michael Steinbach; Anuj Karpatne; Vipin Kuma Introduction to Data Mining 2nd ed, Publisher: Pearson, 2019, Print ISBN: 9780133128901, 0133128903 eText ISBN: 9780134080284, 013408028

[6] M. Kantardzic, Data Mining Concepts, Models, Methods, and Algorithms, 2nd ed, John Wiley & Sons, Inc., Hoboken, New Jersey 2011, ISBN 978-0-470-89045-5 , oBook ISBN: 978-1-118-02914-5, ePDF ISBN: 978-1-118-02912-1, ePub ISBN: 978-1-118-02913-8

[7] Ahishakiye, Emmanuel & Opiyo, Elisha & Wario, Ruth & Niyonzima, Ivan. (2017). A Performance Analysis of Business Intelligence Techniques on Crime Prediction. International Journal of Computer and Information Technology. 06. 84 - 90.

[8] Marlina, Leni & Muslim, Muslim & Siahaan, Andysah Putera Utama. (2016). Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms). International Journal of Emerging Trends & Technology in Computer Science. 38. 380-383. 10.14445/22315381/IJETT-V38P268.

[9] Himani Bhavsar, Mahesh H. Panchal, (2012). A Review on Support Vector Machine for Data Classification, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 1, Issue 10, December 2012, ISSN: 2278 – 1323.

[10] Xiaohu, Wang & Lele, Wang & Nianfeng, Li. (2012). An Application of Decision Tree Based on ID3. Physics Procedia. 25. 1017-1021. 10.1016/j.phpro.2012.03.193.

[11] Hssina, Badr & MERBOUHA, Abdelkarim & Ezzikouri, Hanane & Erritali, Mohammed. (2014). A comparative study of decision tree ID3 and C4.5. (IJACSA) International Journal of Advanced Computer Science and Applications. Special Issue on Advances in Vehicular Ad Hoc Networking and Applications. 10.14569/SpecialIssue.2014.040203.

[12] Kumbhare, Trupti A. and Santosh V. Chobe. "An Overview of Association Rule Mining Algorithms." (2014).

[13] Chengqi Zhang Shichao Zhang, Association Rule Mining Models and Algorithms, ISSN 0302-9743, ISBN 3-540-43533-6 Springer-Verlag Berlin Heidelberg New York, Springer, 2002

[14] Larose, Daniel T. Data mining methods and models, Published by John Wiley & Sons, Inc., Hoboken, New Jersey, 2006, ISBN-13 978-0-471-66656-1 ISBN-10 0-471-66656-4