INTERNATIONAL JOURNAL OF MATHEMATICAL ANALYSIS AND OPTIMIZATION: THEORY AND APPLICATIONS
VOL. 2017. , PP. 139 -153

Classification of Crime Data for Crime Control Using C4.5 and Naïve Bayes Techniques

Obuandike Georgina N. [1*], John Alhasan [2], M. B. Abdullahi [3]

1. *Department of Mathematical Sciences and IT, Federal University Dutsinma, Katsina state, Nigeria.*
2. *Department of Computer Science, Federal University of Technology, Minna, Niger State, Nigeria*
3. *Department of Computer Science, Federal University of Technology, Minna, Niger State, Nigeria*

* Corresponding author: nkoliobunadike@yahoo.com

Article Info
Received: 9 March 2017 Revised: 6 April 2017
Accepted: 6 July 2017 Available online: 21 July 2017

Abstract

The analysis of crime data helps to unravel hidden trends that will aid in better understanding of crime pattern and the nature of those who commit such crimes. It also enables appropriate strategies to be put in place to control such crimes. Literature revealed that C4.5 and Naïve Bayes are effective classification algorithms that have been successfully applied in classification problems. Percentage split or 10 fold cross validation are two approaches to training and testing classifiers. The two approaches were adopted in the training and testing of Naïve Bayes and C4.5 classifiers on crime data collected from selected Nigerian Prisons. In this article, the crime dataset is classified into vulnerable and non-vulnerable for effective crime control strategies. The classification algorithms are applied individually on real crime data and their performance evaluation is analyzed using standard measures such as accuracy, time, Receivers Operating Characteristic (ROC). The classification algorithms are also applied on Breast Cancer and Irish data sets for reliability test. The result showed that C4.5 performed better with higher accuracy on the three dataset against Naïve Bayes. The result also revealed that the two classifiers performed better under percentage split approach compared to 10 fold validation approaches.

Keywords: Data Mining, Crime Analysis, Naïve Bayesian, C4.5

MSC2010: 68T10

1 Introduction

Crime is a societal ill and its cost is usually enormous hence, the need for analysis of crime data to learn the factors that enhance crime and the nature of offenders (Brown, 2003). Many people desire to live and operate in a secure environment;

they want to be sure that their lives and that of their loved ones are secured. The main duty of any Government is to secure the lives and properties of its citizens through relevant policies and strategies. Nigeria is currently having serious security challenges ranging from the Boko Haram attacks in the North-East to armed robbery and kidnapping in other parts of the country. Data mining has gained recognition in crime analysis (Jiawei et al, 2012). It is a field that cuts across many other fields. There are many definitions of data mining according to related literature (Julio and Adem, 2009). The emergence of computing and communication technology has produced a society that extremely depends on information (Witten and Frank, 2000). Development in technology has helped in collection and storage of large amount of data in many organizations' database and these usually contain hidden information. Most of these organizations gather these data for operational purposes after which they are dumped in data repositories or even thrown away or deleted. This type of data when mined can help in discovering of relevant information which can help the organization to increase productivity and can also serve as essential information to the society at large. Data mining has the capability to unravel the information that is usually hidden in such databases. Naisbitt (1986) is of the opinion that we are being choked with data but lack relevant information because the data are not mined to get relevant information. Data mining tools and algorithms are used to find relevant trends and to make necessary predictions and associations in data. Data mining has been successfully applied in virtually all areas of human endeavour which include banking, marketing, manufacturing, telecommunication, e-commerce and education (ZaoHui and Jamie, 2005). Data mining is an intelligent and potent data extraction technique that uses different types of data extraction algorithms. Data analysts explore large data repositories by using these data mining algorithms (Chen et al, 2004; Fayyad and Uthurusamy, 2002)

The analysis of crime data will help in cost reduction and reduction of training time of officer involved in crime control. It will also help in distribution of scarce resources to the appropriate quarters (Megaputer, 2002). The rest of the sections discussed about data mining techniques, data mining process; classification of the crime data using two popular classifiers and discussion of results.

2 Related Works

Many works in literature that performed comparative analysis on classifiers can be grouped into two groups based on the nature and methodology applied (Yang et al., 2005; Seetha et al., 2011 and Tsang et al., 2005). The first group involved comparison of some old methods in order to justify a new method. Second group involved comparing different classifiers qualitatively or quantitatively. The qualitative comparison involves studying some set of classifiers and stating their advantages and drawbacks of the methods as shown in the works of (Jain et al., 2000; Wu et al., 2008 and Kotsiantis, 2007) without carrying out any experiment with physical dataset. The quantitative analysis involves carrying out a comparative analysis using some dataset. This quantitative method of comparison was demonstrated in the work of (Demsar, 2006) in which about 491 research work that performed quantitative tests on classifiers were analyzed. (Huang, 2005) also conducted a quantitative analysis on three different classifiers namely Support Vector Machine (SVM), Decision tree and Naïve Bayes and was of the opinion that Naïve Bayes performs better than decision tree using ROC as performance metric. The work of (Ripley, 1994) also involves comparison of neural network and other classifiers over some dataset. Moreover, quantitative comparative studies of classifiers are usually applied in different area of human endeavour. This is demonstrated in the work of (Tavares, 2008) in the area of bioinformatics, (Cufoglu et al., 2009 and Mico and Oncina, 1998) in area of computer science, (Conrad et al., 2004 and Kuramochi and Karypis, 2005) in the area of medicine and (Berrueta et al., 2007) in the area of chemistry.

In this work, a quantitative analysis that uses Naïve Bayes and C4.5 classifiers on real crime dataset and some selected UCI dataset for viability test on the algorithms performance is conducted. There are other classifiers used in data mining according to literature for classification task these include ID3, SVM, KNN.
Nearest Neighbour (KNN) is a simple algorithm used for classification or regression. The simplicity of this algorithm has made it popular among other classifiers though it is a lazy classifier, can compete effectively among other classifiers. The work of (Neelamegam et al., 2013) enumerates its drawbacks.

ID3: there are many decision tree methods that have been developed over the years among which is ID3 which handles nominal data and it's a greedy classifier (Kalpana and Bansal, 2014; Hongbo, 2010)

Support Vector Machine: There are many versions of SVM classifiers that have been developed to improve classification accuracy; it has been successfully applied in many different areas such as bioinformatics, text classification, and pattern recognition. It uses hyper plane to create an optimal margin classifier thus usually focused on searching for a better strategy of learning optimal separating hyper plane in a large dataset (Tan, et al., 2006; Adewara and Mbata,2016)

3 Classification Techniques

Classification is a technique used to predict an unknown class label using a function. Classification as a method comes in two steps, the first step involves the construction of the classification model (model training) while the second step involve using the model to predict class labels. An instance R of an m-dimensional attribute vector can be represented as $R = (r_1, r_2 \ldots r_m)$ each instance belongs to a class of determined attributes $T_1, T_2 \ldots . T_m$. When an attribute class is discrete value or unordered, it is said to be a categorical or nominal attribute and it serves as the category or fields of the records. The records that are used for the construction of the classification model are usually trainers and are portioned out from the dataset being used for analysis. The training model can be represented as a function $Z = f(r)$ which represents the used fields Z of a given record R (Jiawei et al, 2012).

3.1 C4.5 Classifier

C4.5 is a statistical classifier that is used to create a decision tree. It is carved out from ID3 method to overcome its methodological challenges by pruning the decision tree after construction and handling discrete and continuous dataset. It applies entropy method in choosing the optimal attribute that will effectively divide the training dataset into different classes. It computes entropy (information gain) for all attribute and chooses the attribute with the optimal entropy as test attribute from the dataset (Hongbo, 2010; Kalpana and Bansal, 2014). This is one the major strategy in decision tree methodology which helps in selecting the attribute with the optimal value for the split. Information gain is the sensitive step in the construction of decision tree. It is the information gain that is used in picking the required attribute at every iteration. Let D be dataset $d_1, d_2, \ldots d_n$ with m dimensional attributes $t_1, t_2, \ldots t_m$ and $k_1, k_2, \ldots k_i$ represents class groups. The information gain is given by:

$$\mathrm{Info}(t) = -\sum_{i=1}^{n} P_i \log_2(P_i) \tag{3.1}$$

P_i implies probability that a tuple in D belongs to class k_i.
The attributes t can be used to partition the dataset D into $\{d_1, d_2, \ldots d_k\}$ where $t_i \in d_i$ can be represented as

$$\mathrm{info}_t\,(D) = \sum_{i=1}^{k} \frac{|d_i|}{|d|} \times \mathrm{info}\,(d_i) \tag{3.2}$$

Where $\frac{d_i}{d}$ implies weight to ith partition, $\mathrm{info}_t(d)$ is the information gain to be used for classification.

$$\mathrm{Gain}\,(t) = \mathrm{info}\,(D) - \mathrm{info}_t(D) \tag{3.3}$$

Thus the optimal attribute is chosen as test attribute. C4.5 is an improvement of ID3 classifier, it employs gain ratio, a type of information gain to improve ID3 methodology. It uses a normalization method called split information usually define as:

$$\mathrm{Sinfo}_t(D) = \sum_{i=1}^{k} \frac{|d_i|}{|d|} \times \log_2\{\frac{|d_i|}{|d|}\} \tag{3.4}$$

Calculating the gain ratio as:

$$\mathrm{ganRatio}(t) = \frac{\mathrm{gain(t)}}{\mathrm{Sinfo(t)}} \tag{3.5}$$

At each point in the tree C4.5 algorithm usually pick an attribute that gave the best split of the dataset. The attribute with the highest normalized value is chosen for the split and it is placed at the root of the tree. C4.5 is a supervised learning method that is simple and easy to implement. It divides dataset into portions with different characteristics. The last leave of the tree usually depicts predictions while the in between nodes depicts various test on the attributes (from the root node to the leaf node).

3.2 Naïve Bayes Classifier

Naïve Bayes classifier is a probability based classifier and has proved its effectiveness in many areas where it has been applied. It is fast and easy to use which made it popular in data mining field. Though usually criticized for its attribute independent assumptions but it still competes favourably with other higher classifiers. Naïve Bayes calculates the probability value and selects the class with the highest probability (Taheri et al, 2014). It is represented mathematically as shown in Equation 3.6 and Equation 3.7. The algorithm is also represented in Figure 1.

$$P(K_i \cap Y) = \frac{P(Y \cap K_i)P(K_i)}{P(Y)} \tag{3.6}$$

For a database with high dimension the computational cost is usually high thus the application of Naïve Bayes.

$$P(Y \cap K_i) = \prod_{k=1}^{n} P(Y_k \cap K_i) \tag{3.7}$$

Naïve Bayes Algorithm

Input: Data attributes and the class of the instances

Output: Class labels

Method:

Compute the posterior value for each attribute against the class using $P(K/Y) = \frac{P(K/Y_i)P(Y_i)}{P(Y)}$

Compute the value before the existing class: $P(Y/K_i)P(K_i)$

Multiply the results from 3 and 4 for all the classes and choose the highest value as the classification: $P(Y/K_i) = \prod_{q=1}^{n} P(Y_{q/}K_i)$

Output the result

Figure 1: Naïve Bayes Algorithm (Taheri et al, 2014)

4 Methodology

The crime dataset used for this research was collected from selected Nigerian prisons. It has 1733 records with six attributes including the class attribute. A sample of the crime data set and the attributes is as shown in Table 1. Before the application of the classification algorithms, an exploratory analysis was done to get a better insight into the dataset before analysis. This research was implemented using a software called Waikato Environment For Knowledge Analysis (WEKA). It is a free complete data mining software written in Java that contains all necessary tools for data mining ranging from data pre-processing to analysis. The data was analyzed using two approaches firstly, the crime data set was divided into two 66% was used for training and 34% was used for testing the data set. Secondly, 10 fold cross validation was also applied. The results from the two approaches were evaluated using standard performance measures.

Table 1: Crime data set and their attributes including the class attribute

Offence	Age	Sex	Edu-Qualification	Occupation	Class
low	early	M	high	Self Employed	non vulnerable
high	early	M	average	Self Employed	vulnerable
low	early	M	average	Self Employed	non vulnerable
low	middle	M	high	employed	non vulnerable
low	late	M	NONE	employed	non vulnerable
low	middle	M	NONE	Self Employed	non vulnerable
low	middle	F	NONE	unemployed	non vulnerable
low	early	M	low	Self Employed	vulnerable
low	late	M	NONE	unemployed	non vulnerable

WEKA is a machine learning software that has gained recognition in data mining because it implements many different data mining algorithms and also has potent tools for data pre-processing and visualization. It is an open source and accepts its data in Attribute Related File Format (ARFF). The sample of the converted ARFF file for this work is shown in figure 2.

```
@relation 'Formated Prisons2-weka.filters.unsupervised.attribute.Remove-R1-2,11-19-
weka.filters.unsupervised.attribute.Remove-R7-weka.filters.unsupervised.attribute.Remove-
R3-weka.filters.unsupervised.attribute.Remove-R3'

@attribute Offence {low,high,average}
@attribute Age {early,late,middle}
@attribute Sex {M,F,'M '}
@attribute Edu-Qualification {low,average,NONE}
@attribute Occupation {'Self Employed',unemployed,employed}
@attribute Class {vulnerable,'non vulnerable'}

@data
low,early,M,low,'Self Employed',vulnerable
high,early,M,average,'Self Employed',vulnerable
low,early,M,average,'Self Employed',vulnerable
average,early,M,low,'Self Employed',vulnerable
average,early,M,low,'Self Employed',vulnerable
high,early,M,average,'Self Employed',vulnerable
average,early,M,low,'Self Employed',vulnerable
low,early,M,low,'Self Employed',vulnerable
high,early,M,low,'Self Employed',vulnerable
high,early,M,low,'Self Employed',vulnerable
low,early,M,low,'Self Employed',vulnerable
low,early,M,average,unemployed,vulnerable
average,late,M,NONE,employed,'non vulnerable'
low,early,M,average,'Self Employed',vulnerable
average,early,M,average,'Self Employed',vulnerable
```

Figure 2: A Sample ARFF for the Crime dataset

5 Performance Measures

The performance of the classifiers on the crime dataset were analyzed using standard metric for evaluating classifiers performance such as sensitivity, relevance, specificity, kappa statistics, area curve, time and accuracy.

Sensitivity (TP Rate): It is a statistics that shows the records that are correctly labeled by the classifier. It can be defined as: $\text{Sensitivity} = \text{TP}/\text{N}$

Specificity (FP Rate): It is simply a report of instances incorrectly labeled as correct instances; it can be defined as: $\text{Specificity} = \text{FP}/\text{N}$

Precision: Simply measures exact relevant data retrieved. High precision means the model returns more relevant data than irrelevant data. $\text{Precision} = \text{TP} / \text{TP} + \text{FP}$

Kappa: measures the relationship between classified instances and true classes. It usually lies between $[0, 1]$, the value of 1 means perfect relationship while 0 means random guessing.

Accuracy: this shows the percentage of correctly classified instances in each classification model

Time: Implies time taken to perform the classification (Milan and Sunila, 2011; Hong etal, 2006)

Where:

TN implies True Negative: that is number of correct predictions that an offender is vulnerable

FP implies False Positive that is number of wrong predictions that an offender is non vulnerable

FN implies false Negative that is no of wrong predictions that an offender is vulnerable

TP implies True Negative that is number of correct predictions that an offender is non vulnerable

6 Experimental Results

When preparing data for data mining seeing the data pictorially provides insight into what is happening and this insight can help improve model building. The data mining tool chosen for this work has the features for exploratory data analysis. The relative densities of the various attributes in the data set are as shown in figure 3. The visualization of relationships between important attributes is as shown in Figure 4 and Figure 5.

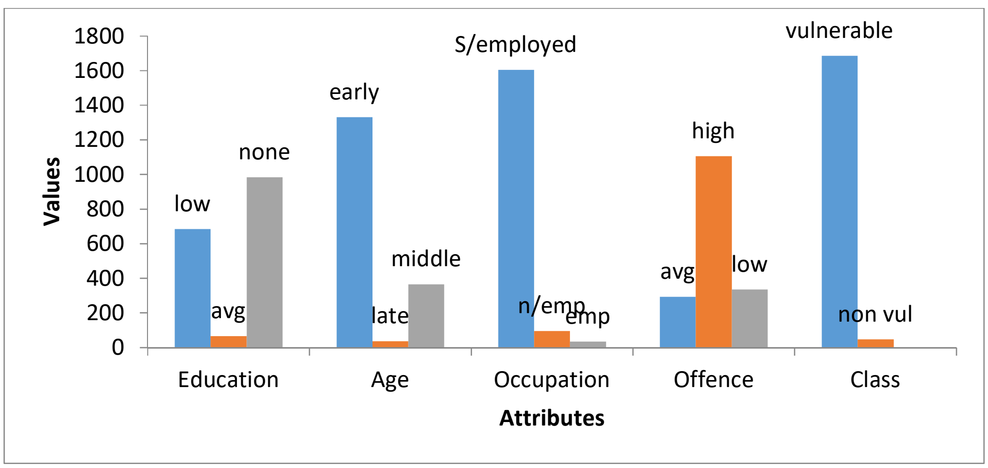

Figure 3: Densities of Attributes in the dataset

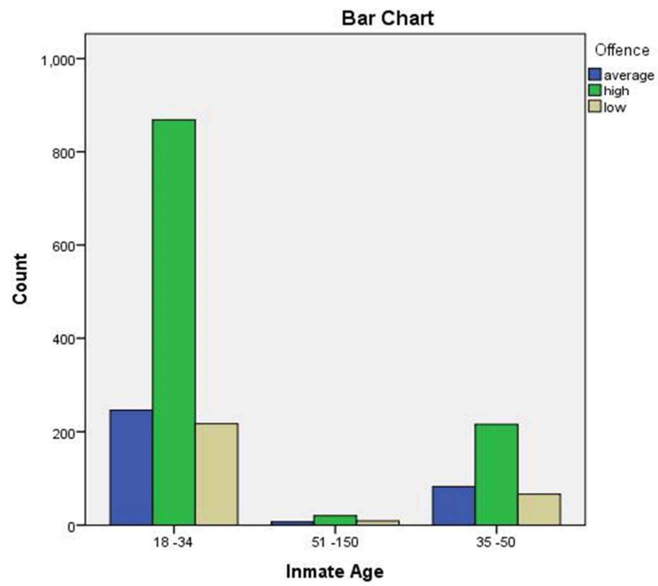

Figure 4: Association of Age and Offence

The visualization of the relationship between the age and offence reveals that majority of the offenders are within the ages of 18 -34 (early age) and commit high crime

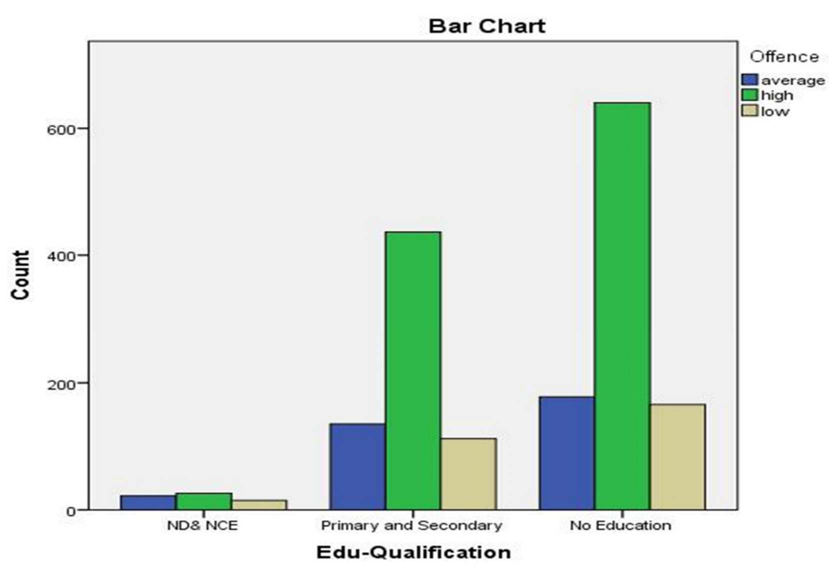

Figure 5: Association of Educational Qualification and Offence

The visualization of the offence versus educational qualification reveals that majority of the offenders have low education qualification (primary and secondary) or no education qualification at all and that they commit high crime and are vulnerable groups.

6.1 Classification Result using percentage split

This is a situation where by the crime dataset used were divided into two and 66% of it was used for training the classifiers while 34% was used for testing the classifiers. The result obtained using the percentage split is as shown in Table 2

Table 2: Result of classification of crime dataset using percentage split approach

	Evaluation Metrics							
classifiers	Time	Accuracy	TP Rate	FP Rate	Kappa	Precision	Recall	ROC
C4.5	0.08	97.5	0.975	0.023	0.949	0.975	0.975	0.993
NB	0.02	95.67	0.957	0.04	0.913	0.958	0.957	0.99

The result above revealed that the C4.5 classifier has better accuracy of 97.5 in comparison to the accuracy of 95.67 from Naïve Bayes. C4.5 though took more time of 0.08 seconds to build the model compare to 0.02 seconds taken by the Naïve Bayesian still handles the data better. In terms of classifier performance using the ROC curve the C4.5 classifier performed comparably well against Naïve Bayes on the dataset.

ROC curve is used to visualize classifiers performance. It is usually plotted using sensitivity at the y axis and specificity at the x axis. If the area under the curve is 1, it indicates perfect prediction while 0.5 implies random guess. The areas under the curve for the Naïve Bayesian and C4.5 classifiers are close to 1 which indicates the classifiers performed well.

6.2 Classification Result from 10 Fold Cross Validation

Table 3 showed the classification results from the two classifiers Naïve Bayes (NB) and C4.5 using 10 Fold cross validation approach. The result showed that C4.5 has higher accuracy of 97.06 compared to 93.54 given by Naïve Bayes. The result also showed that the two classifiers performed better under percentage split approach than with the 10 fold cross validation.

Table 3: Result of classification of crime dataset using 10 fold cross validation approach

	Evaluation Metrics							
classifiers	Time	Accuracy	TP Rate	FP Rate	Kappa	Precision	Recall	ROC
C4.5	0.09	97.06	0.971	0.027	0.940	0.971	0.971	0.971
NB	0.01	93.54	0.935	0.067	0.870	0.935	0.935	0.989

6.3 Viability Test Result for the two classifiers using Breast Cancer and Irish dataset

A viability test was performed using the two algorithms on two UCI dataset (Breast cancer and Irish data set) to establish the performance of the two approaches using the two approaches of percentage split and 10 fold cross validation. The result of classification using percentage split for breast cancer and Irish data set is as shown in Table 4 and Table 5.

Table 4: Result of classification of breast cancer dataset using percentage split approach

	Evaluation Metrics							
classifiers	Time	Accuracy	TP Rate	FP Rate	Kappa	Precision	Recall	ROC
C4.5	0	75.87	0.759	0.523	0.2899	0.76	0.759	0.639
NB	0	75.18	0.752	0.404	0.3693	0.74	0.752	0.76

Table 5: Result of classification of Irish dataset using percentage split approach

	Evaluation Metrics							
classifiers	Time	Accuracy	TP Rate	FP Rate	Kappa	Precision	Recall	ROC
C4.5	0.05	98	0.98	0.01	0.97	0.98	0.98	0.993
NB	0.01	96	0.96	0.02	0.94	0.96	0.96	0.995

It was observed from Table 4 and Table 5 that C4.5 algorithm performed better than Naïve Bayes algorithm using percentage split on Breast Cancer and Irish dataset. The classification result of Breast Cancer and Irish dataset using 10 fold cross validation approach is as shown in Table 6 and Table 7.

Table 6: Result of classification of Breast Cancer dataset using percentage 10 fold cross validation approach

	Evaluation Metrics							
classifiers	Time	Accuracy	TP Rate	FP Rate	Kappa	Precision	Recall	ROC
C4.5	0.11	75.53	0.755	0.524	0.2826	0.752	0.755	0.584
NB	0.02	71.68	0.717	0.446	0.2857	0.704	0.717	0.701

Table 7: Result of classification of Irish dataset using percentage 10 fold cross validation approach

	Evaluation Metrics							
classifiers	Time	Accuracy	TP Rate	FP Rate	Kappa	Precision	Recall	ROC
C4.5	0.05	96	0.96	0.02	0.94	0.96	0.96	0.968
NB	0	96	0.96	0.02	0.94	0.96	0.96	0.994

The classification results from the two approaches using breast cancer and Irish dataset reveals also that the classifiers performs better under percentage split approach compared to the 10 fold cross validation approach.

7 Conclusion

Data mining has the capability that makes it simple convenient and suitable for data extraction from large databases. It employs different mining algorithms for its work. Many agencies gather data for its operational purposes, such data can be mined to discover some relevant patterns that can aid in decision making. The analysis of crime data will help to unravel crime pattern and nature of those who commits such crime so that appropriate strategies and rules will be put in place to control such crimes.

To achieve the objective of this research, two different approaches were adopted in training and testing the classifiers on the crime dataset. The two approaches adopted are percentage split approach which divides the dataset into 66% for training and 34% percent for testing the trained classifier. The other method is 10 fold cross validation, this method usually divides the dataset into k-folds and trains the model with k-1 fold and test the trained model with remaining K fold. It usually obtained k different results and takes the average to obtain the model accuracy. The classification algorithms were evaluated using eight performance measures such as time, TP rate, FP rate, kappa, precision, recall, accuracy and ROC.

The experiments using percentage split and 10 fold cross validation reveals that C4.5 gave higher accuracy than Naïve Bayes under the two approaches. The result also reveals that the classifiers performed better under percentage split. A viability test was also conducted using two UCI dataset namely: Breast Cancer and Irish dataset to establish the performance of the algorithms using percentage split and 10 fold cross validation. The result obtained from the viability test also revealed that C4.5 gave higher accuracy than Naïve Bayes on Breast Cancer and the Irish dataset. The viability result equally showed that the two classifiers performed better under percentage split approach.

Acknowledgements

The authors acknowledge all that have contributed in one way or the other in making this research a success. We acknowledge particularly the assessors for their constructive criticisms and the management of International Journal of Mathematical Analysis and Optimization for accepting to publish this research work in their reputable journal.

Competing financial interests

No competing financial interests.

References

[1] Adewara A. J. & Mbata A. Optimization of exponentially weighted moving average statistics using empirical Bayesian factor. International Journal of Mathematical Analysis and Optimization, 3, 83 – 94(2016)

[2]Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. Supervised pattern recognition in food analysis. Journal of Chromatography A, 1158(1), 196-214 (2007).

[3] Brown, D. E. The regional crime analysis program (RECAP). A framework for mining data to catch criminals. In Systems, Man, and Cybernetics, 3(1), 2848-2853. (2003).

[4] Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., & Chau, M. Crime Data Mining: A General Framework and Some Examples. Computer, 37(4), 50-56 (2004).

[5] Conrad, C., Erfle, H., Warnat, P., Daigle, N., Lörch, T., Ellenberg, J., ... & Eils, R. Automatic identification of subcellular phenotypes on human cell arrays. Genome research, 14(6), 1130-1136 (2004).

[6] Cufoglu, A., Lohi, M., & Madani, K. A comparative study of selected classifiers with classification accuracy in user profiling. In Computer Science and Information Engineering, 3, 708-712 (2009).

[7] Demsar, J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 7(1), 1-30 (2006).

[8] Fayyad, U.M., & Uthurusamy, R. Evolving Data Mining into Solutions for Insights, Communications of ACM, 45(8), 28-31(2002).

[9] Huang, J., & Ling, C. X. Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions Knowledge and Data Engineering 17, 299-310 (2005).

[10] Hong, H., Jiuyong, L., & Ashley P. A Comparative Study of Classification Methods for Microarray Data Analysis", published in CRPIT, 61(1), (2006).

[11] Hongbo, D. Data Mining Techniques and Applications-an Introduction, Cenage Learning EMEA. (2010).
[12] Jain, A. K., Duin, R. P. W., & Mao, J. Statistical pattern recognition: A review. IEEE Transactions on pattern analysis and machine intelligence, 22(1), 4-37 (2000).
[13] Jiawei, H., Micheline, K., & Jian P. Data mining: Concept and Techniques, 3rd edition, New York: Elsevier, (2012).
[14] Julio, P., and Adem, K. Data Mining and Knowledge Discovery in Real Life Applications.Vienna, Austria:I-Tech, (2009).
[15] Kalpana, R., & Bansal, K. L. A Comparative Study of Data Mining Tools, International Journal of Advanced Research in Computer Science and Software Engineering, 4, (2014).
[16] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. Supervised machine learning: A review of classification techniques. Emerging Application in Computer Engineering, 160(1), 3-24 (2007).
[17] Kuramochi, M., & Karypis, G. Gene classification using expression profiles: a feasibility study. International Journal on Artificial Intelligence Tools, 14, 641-660 (2005).
[18] Kurgan, L., & Musilek, P. A survey of knowledge discovery and data mining process models. The Knowledge Engineering Review, 21(1), 1 – 24 (2006).
[19] Megaputer Intelligence, Inc. Crime Pattern Analysis: Megaputer Case Study.http://www.elon.edu/facstaff/mconklin/cis230/cases/crime_pattern_case.pdf (2002).
[20] Micó, L., & Oncina, J. Comparison of fast nearest neighbour classifiers for handwritten character recognition. Pattern Recognition Letters, 19(3), 351-356 (1998).
[21] Milan, K., & Sunila, G. Comparative Study of Data Mining Classification Methods in cardiovascular Disease Prediction. International Journal of Computer Science and Technology, 2 (2), 304-308 (2011).
[22] Naisbitt, J. (1986). Megatrends (6th ed.). New York, Warner Books.
[23] Neelamegam, S., & Ramaraj, E. Classification algorithm in data mining: An overview. International Journal of P2P Network Trends and Technology (IJPTT), 4(8), 369-374 (2013).
[24] Ripley, B. D. Neural networks and related methods for classification. Journal of the Royal Statistical Society. Series B (Methodological), 409-456 (1994).
[25] Seetha, H., & Saravanan, R. On improving the generalization of SVM classifier. In Computer Networks and Intelligent Computing (pp. 11-20). Springer Berlin Heidelberg (2011).
[26] Taheri, S., Yearwood, J., Mammadov, M., & Seifollahi, S. Attribute weighted Naive Bayes classifier using a local optimization. Neural Computing and Applications, 24(5), 995-1002 (2014).
[27] Tan, P. N., Steinbach, M., & Kumar, V. Introduction to Data Mining, Indina: Pearson Addison Wesley, (2006).
[28] Tavares, L. G., Lopes, H. S., & Lima, C. R. E. A comparative study of machine learning methods for detecting promoters in bacterial DNA

sequences. In International Conference on Intelligent Computing, 959-966 (2008).

[29] Tsang, I. W., Kwok, J. T., & Cheung, P. M. Core vector machines: Fast SVM training on very large data sets. Journal of Machine Learning Research, 6(4), 363-392 (2005).

[30] Witten, I., & Frank, E. Data mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco: Morgan Kaufmann publishers, (2000).

[31] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z. H. Top 10 algorithms in data mining. Knowledge and information systems, 14(1), 1-37 (2008).

[32] Yang, J., Frangi, A. F., Yang, J. Y., Zhang, D., & Jin, Z. KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. IEEE Transactions on pattern analysis and machine intelligence, 27(2), 230-244 (2005).

[33] ZhaoHui, T., & Jamie, M. Data Mining with SQL Server 2005. Indianapolis, Indiana: Wiley Publishing (2005).