International Journal of Computer and Information Technology (ISSN: 2279 – 0764)
Volume 06 – Issue 03, May 2017

Crime Prediction Using Decision Tree (J48) Classification Algorithm

Emmanuel Ahishakiye
School of Computing and Informatics,
University of Nairobi, P.O. Box 30197 – 00100,
GPO Nairobi-Kenya
Email: ahishema [AT] gmail.com

Elisha Opiyo Omulo
School of Computing and Informatics,
University of Nairobi, P.O. Box 30197 – 00100,
GPO Nairobi-Kenya

Danison Taremwa
Department of Computer science, Kyambogo University,
P.O. Box 1, Kyambogo, Kampala – Uganda

Ivan Niyonzima
Department of ICT, Faculty of science,
Kabale University, P.O. Box 317
Kabale – Uganda

Abstract—**There had been an enormous increase in the crime in the recent past. Crimes are a common social problem affecting the quality of life and the economic growth of a society. With the increase of crimes, law enforcement agencies are continuing to demand advanced systems and new approaches to improve crime analytics and better protect their communities. Decision tree (J48) applied in the context of law enforcement and intelligence analysis holds the promise of alleviating such problem. Data mining is a way to extract knowledge out of usually large data sets; in other words it is an approach to discover hidden relationships among data by using artificial intelligence methods of which decision tree (J48) is inclusive. The wide range of machine learning applications has made it an important field of research. Criminology is one of the most important fields for applying data mining. Criminology is a process that aims to identify crime characteristics. This study considered the development of crime prediction prototype model using decision tree (J48) algorithm because it has been considered as the most efficient machine learning algorithm for prediction of crime data as described in the related literature. From the experimental results, J48 algorithm predicted the unknown category of crime data to the accuracy of 94.25287% which is fair enough for the system to be relied on for prediction of future crimes.**

Keywords: crime prediction; machine learning; decision tree; J48; artificial intelligence; Classification Algorithms.

I. INTRODUCTION

Crimes are a common social problem affecting the quality of life and the economic growth of a society [1]. It is considered an essential factor that determines whether or not people move to a new city and what places should be avoided when they travel [2]. The effects of crime on society include feelings of fear that disrupt the population's sense of unity, the breakdown of social associations due to habitual avoidance of certain places, an unwillingness to go out at night and damage to the image of the community. The perception of a community as crime ridden can deter people from going there and induce residents to move away. This causes damage to the economy. Crime affects the economy by placing a financial burden on taxpayers and governments because of increased needs for police, courts and corrections facilities, as well as intangible costs including psychological trauma and reduced quality of life for crime victims.

Today, a high number of crimes are causing a lot of problems in many different countries. In fact, scientists are spending time studying crime and criminal behaviors in order to understand the characteristics of crime and to discover crime patterns. Dealing with crime data is very challenging as the size of crime data grows very fast, so it can cause storage and analysis problems. In particular, issues arise as to how to choose accurate techniques for analyzing data due to the inconsistency and inadequacy of these kinds of data. These issues motivate scientists to conduct research on these kinds of data to enhance crime data analysis. Dealing with crime data is very challenging as the size of crime data grows very fast, so it can cause storage and analysis problems. In particular, issues arise as to how to choose accurate techniques for analyzing data due to the inconsistency and inadequacy of these kinds of data. These issues motivate scientists to conduct research on these kinds of data to enhance crime data analysis [3]. The objective of this research is to apply suitable machine learning algorithm on crime data to predict the likelihood of a county having low, medium or high violent crimes.

II. LITERATURE SURVEY

A. Criminology and Crime Analysis

Criminology is an area that focuses on the scientific study of crime and criminal behavior and law enforcement and is a process that aims to identify crime characteristics [4]. It is one of the most important fields where the application of data mining techniques can produce important results. Crime analysis, a part of criminology, is a task that includes exploring and detecting crimes and their relationships with criminals. The high volume of crime datasets and also the complexity of relationships between these kinds of data have made criminology an appropriate field for applying data mining techniques. Identifying crime characteristics is the first step for developing further analysis. The knowledge that is gained from data mining approaches is a very useful tool which can help and support police forces [5]. According to [6], solving crimes is a complex task that requires human intelligence and experience and data mining is a technique that can assist Law Enforcement Agencies with crime detection problems. The idea here is to try to capture years of human experience into computer models via data mining.

B. Why Crime Is Predictable

There is a strong body of evidence to support the theory that crime is predictable (in the statistical sense) mainly because criminals tend to operate in their comfort zone [7]. That is, they tend to commit the type of crimes that they have committed successfully in the past, generally close to the same time and location. Although this is not universally true, it occurs with sufficient frequency to make these methods work reasonably well. There are major theories of criminal behavior, such as routine activity theory, rational choice theory, and crime pattern theory. These theories are consolidated into what is referred to as a blended theory.

C. Review of Classification Algorithms

Classification algorithms that are mostly used in predictions basing on historical data. Classification is a class prediction technique, which is supervised in nature. This technique possesses the ability to predict the label for classes, provided that sufficient numbers of training examples are available. There is a variety of classification algorithms available, including Support vector machines, k Nearest Neighbors, weighted voting and Artificial Neural Networks. All these techniques can be applied to a dataset for discovering set of models to predict the unknown class label. In classification, the dataset is divided into two sets, namely the training set (dependent set) and a test set (independent set). The machine learning algorithm initially runs on the training set, than later the predicting model is applied on the test set. The following are classification algorithms that are used in crime predictions.

1) Decision Tree classifier (DT)

Decision tree learning uses a decision tree as a predictive model which maps observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modeling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a finite set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data (but the resulting classification tree can be an input for decision making). To generate a decision tree, the C4.5 [8] algorithm is used, which is an extension of Quinlan's earlier ID3 algorithm. To construct the tree, entropy measure is used in the determination of nodes. Since the attributes with higher the entropy cause more uncertainty in outcome, they were selected in order of entropy.

A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. See the examples illustrated in the figure for spaces that have and have not been partitioned using recursive partitioning, or recursive binary splitting. The recursion is completed when the subset at a node has all the same value of the target variable, or when splitting no longer adds value to the predictions. This process of top-down induction of decision trees (TDIDT) is an example of a greedy algorithm, and it is by far the most common strategy for learning decision trees from data [9].

In data mining, decision trees can be described also as the combination of mathematical and computational techniques to aid the description, categorization and generalization of a given set of data. Data comes in records of the form:

$$(\mathbf{x}, Y) = (x_1, x_2, x_3, \ldots, x_k, Y)$$

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalize. The vector x is composed of the input variables, x_1, x_2, x_3 etc., that are used for that task.

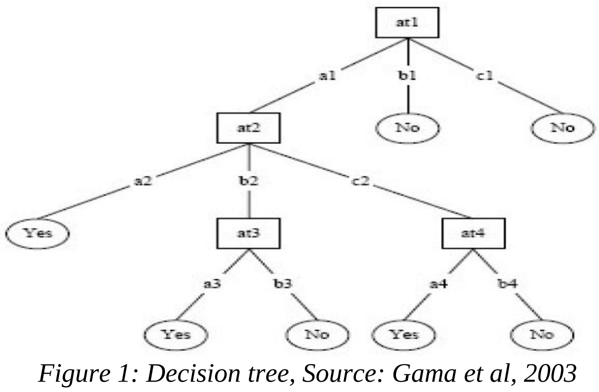

Figure 1: Decision tree, Source: Gama et al, 2003

2) Multilayered Perceptron (MLP)

A Multilayer Perceptron is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. MLP utilizes a supervised learning technique called back propagation for training the network.

The Multilayer Perceptron consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes and is thus considered a deep neural network [10]. Each layer is made up of units. The inputs to the network correspond to the attributes measured for each training tuple. The inputs are fed simultaneously into the units making up the input layer. These inputs pass through the input layer and are then weighted and fed simultaneously to a second layer of "neuron like" units, known as a hidden layer. The outputs of the hidden layer units can be input to another hidden layer, and so on. The number of hidden layers is arbitrary, although in practice, usually only one is used. To improve the classification accuracy we should reduce the training time of neural network and reduce the number of input units of the network [11].

A multi-layer neural network consists of large number of units (neurons) joined together in a pattern of connections. The multilayer Perceptron consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes and is thus considered a deep neural network. Since an MLP is a Fully Connected Network, each node in one layer connects with a certain weight w_{ij} to every node in the following layer. Some people do not include the input layer when counting the number of layers and there is disagreement about whether w_{ij} should be interpreted as the weight from i to j or the other way around. A multilayer feed-forward neural network consists of an input layer, one or more hidden layers, and an output layer. An example of a multilayer feed-forward network is shown in Fig 2.

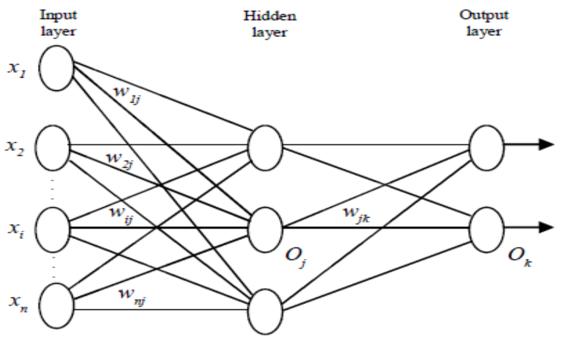

Figure 2: A multilayer feed-forward network. Source: Rohit, 2012

According to [12], each layer is made up of units. The inputs to the network correspond to the attributes measured for each training tuple. The inputs are fed simultaneously into the units making up the input layer. These inputs pass through the input layer and are then weighted and fed simultaneously to a second layer of "neuron like" units, known as a hidden layer. The outputs of the hidden layer units can be input to another hidden layer, and so on. The number of hidden layers is arbitrary, although in practice, usually only one is used. At the core, back propagation is simply an efficient and exact method for calculating all the derivatives of a single target quantity (such as pattern classification error) with respect to a large set of input quantities (such as the parameters or weights in a classification rule). To improve the classification accuracy we should reduce the training time of neural network and reduce the number of input units of the network.

3) Naive Bayes classifiers

Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes has been studied extensively since the 1950s. It was introduced under a different name into the text retrieval community in the early 1960s, and remains a popular (baseline) method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines [13]. It also finds application in automatic medical diagnosis.

Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers. Naive Bayes models are known under a variety of names, including simple Bayes and independence Bayes. All these names reference the use of Bayes' theorem in the classifier's decision rule, but naive Bayes is not (necessarily) a Bayesian method [14].

4) Support Vector Machines

Support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

When data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The clustering algorithm which provides an improvement to the support vector machines is called support vector clustering and is often used in industrial applications either when data are not labeled or when only some data are labeled as a preprocessing for a classification pass [15].

D. Performance analysis of classification algorithms on crime prediction

[16] Analyzed crime data using decision tree and Naïve Bayes algorithms, the accuracy was 83.9519% and 70.8124% respectively. Hence he concluded that decision tree performs better than Naïve Bayes. [17] Did a performance analysis of decision tree (J48), Naïve Bayes, Multilayer Perceptron and support vector machine on crime data; and found that the performance accuracy was 100%, 89.9425%, 100% and 93.6782% respectively with execution time of 0.06sec, 0.14sec, 9.26sec and 0.66sec respectively. Hence decision tree performed better both in performance and in execution time. Therefore, the researcher will use decision tree algorithm (J48) in the proposed BI system because it performs accurately with little time.

E. Conceptual Design of the proposed system

From the previous work done in this field and literature studied and cited above, this study identifies the most appropriate approach. These approaches form the building blocks of a conceptual model used for this research as shown in the figure below.

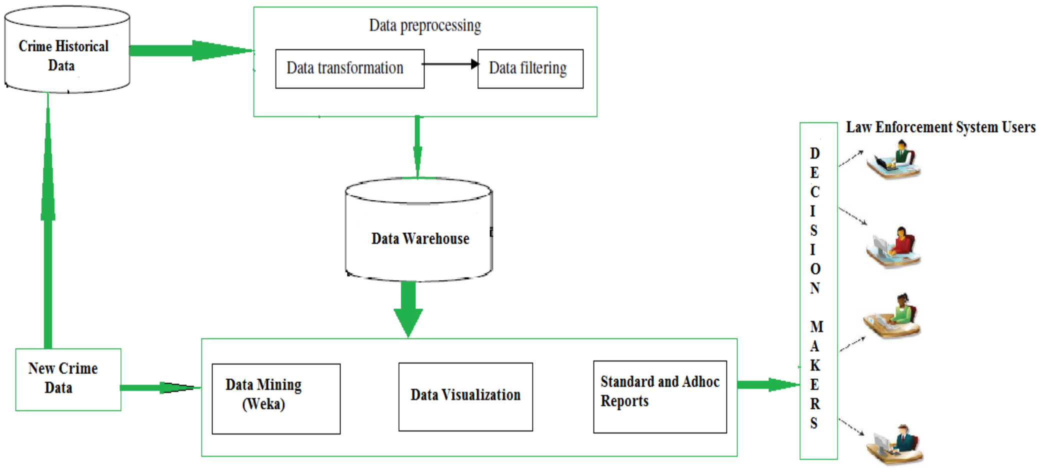

Figure 3: Block diagram of the proposed crime predictive System

III. FRAMEWORK AND METHODOLOGY

Spiral model methodology was used in the system specification, system design and implementation.

A. Overview of spiral model

The spiral model methodology is a systems development lifecycle model which combines the features of the Prototyping Model and the Waterfall Model and has detailed process for specifying, designing, and implementing prototypes [18]. The spiral model is favored for large, expensive and complicated projects [19].

B. The Machine Techniques Used

The machine learning model considered in our study was based on supervised learning (classification) techniques given that labeled training data was available. Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. Also decision tree classifier was applied in crime prediction. Our methodology consists of data collection, data-preprocessing, building classification model using training data and evaluation of the generated models using test data. Trained and tested model was then used to score incoming data.

C. Sources of Data

The dataset used for this study is real and authentic. The dataset was acquired from UCI machine learning repository website. The title of the dataset is 'Crime and Communities'.

This dataset contains a total number of 128 attributes and 1994 instances. All data provided in this dataset is numeric and normalized. The complete details of all 128 attributes can be acquired from the UCI machine learning repository website [20].

D. Attribute selection

The objective of my analysis (crime prediction) did not require all the variables recorded hence there was need for data preparation, reduction and pre-processing. Data reduction is performed by selecting the most informative attributes in a dataset, while attempting to lose no critical information for classification. There was need for removal of the variables which I did not need. From the 128 attributes only 12 were of use for the analysis. There are different methods available for attribute or feature selection but manual method is usually chosen or attribute selection based on human understanding of data set. When dealing with a large number of attributes it is practical to use human knowledge to make decisions on the attributes and also taken in account that only those attributes are chosen which do not contain any missing values.

E. Variables used in this study

State, population, MedIncome **(**Median household income**),** MedFamInc (Median family income (differs from household income for non-family households)), PerCapInc (Per capita income), NumUnderPov (Number of people under the poverty level), PctLess9thGrade (Percentage of people 25 and over with less than a 9th grade education), PctNotHSGrad (Percentage of people 25 and over that are not high school graduates), PctBSorMore (Percentage of people 25 and over with a bachelor's degree or higher education), PctUnemployed (Percentage of people 16 and over, in the labor force, and unemployed), PctEmploy (Percentage of people 16 and over who are employed), ViolentCrimesPerPop (Total number of violent crimes per 100K population), Crime Category (Crime categorization in to three categories, namely). The new added nominal attribute have three values, which are 'Low', 'Medium', and 'High'. If the value in 'Violent Crimes Per Pop' is less than 25 percent than the value of 'Crime Category' is 'Low', If the value in 'Violent Crimes Per Pop' is equal to or greater than 25 percent and less than 40 percent, than the value of 'Crime Category' is 'Medium', If the value in 'Violent Crimes Per Pop' is equal to or greater than 40 percent than the value of 'Crime Category' is 'High'.

F. Modeling Techniques and Tools Used

The model considered in our study was based on supervised learning (classification) technique. The software tool used was WEKA an open-source and free software used for knowledge analysis and downloadable from the internet and used under the GNU license. WEKA implements different machine learning algorithms. The presentation of results and the development of the prototype were done using JAVA while the data will be stored in JavaDB. This Java-based version (Weka 3.8.0) is used in many different application areas, in particular for educational purposes and research.

Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. All techniques of Weka's software are predicated on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported).

IV. RESULTS AND DISCUSSION

This describes the design process and implementation of the software prototype that was built for the purpose of experimentation in this study. The implementation of the system in terms of the data set used, the programming strategies selected and the testing process is outlined.

A. Training data set

To produce the model a training data was used, we used a data set with known output values and use this data set to build our model as in Fig.2. However, this type of model takes an entire training set and divide it into two parts, i.e about 80% of the data is taken and put into our training set, which we use to create the model; then the remaining data set is put into a test data set, which we use immediately after creating the model to test the accuracy of our model.

```
=== Summary ===

Correctly Classified Instances         696              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances              696

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     MEDIUM
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     LOW
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     HIGH
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000

=== Confusion Matrix ===

   a   b   c   <-- classified as
  93   0   0 |   a = MEDIUM
   0 466   0 |   b = LOW
   0   0 137 |   c = HIGH
```

Figure 4: Classification of training data using Decision Tree (J48)

B. Test data set

The test data was created to control over fitting, after the model is created it is tested to ensure that the accuracy of the model built does not decrease with the test set as in Fig.6. This ensures that our model will accurately predict future unknown values.

```
=== Summary ===

Correctly Classified Instances          68              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances               68
Ignored Class Unknown Instances                  9

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    0.972     0.600     MEDIUM
                 1.000    0.000    1.000      1.000    1.000      1.000    0.919     0.902     LOW
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     HIGH
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000    0.943     0.898

=== Confusion Matrix ===

  a  b  c   <-- classified as
  6  0  0 |  a = MEDIUM
  0 46  0 |  b = LOW
  0  0 16 |  c = HIGH
```

Figure 5: Classification of test data using Decision Tree (J48)

C. Prediction of violent crimes

After training and testing our model in fig 1 and 2 respectively, data of unknown crime category was then fed into the system for prediction. The predicted output of a given city is predicted as low, medium or high. Out of 174 datasets supplied into the system, 164 were correctly predicted and only ten were incorrectly predicted. The percentage of incorrectly predicted datasets is 94.25287% as shown in fig.4 below. This percentage is fair enough for the system to be entirely depended on by the law enforcement agencies.

```
Expected values: LOW, Predicted values: LOW
Expected values: MEDIUM, Predicted values: MEDIUM
Expected values: MEDIUM, Predicted values: MEDIUM
Expected values: MEDIUM, Predicted values: LOW
Expected values: LOW, Predicted values: LOW
Expected values: LOW, Predicted values: LOW
Expected values: MEDIUM, Predicted values: LOW
Expected values: LOW, Predicted values: LOW
Expected values: LOW, Predicted values: LOW
Expected values: LOW, Predicted values: LOW
Expected values: MEDIUM, Predicted values: HIGH
Expected values: LOW, Predicted values: LOW
Expected values: LOW, Predicted values: LOW
Expected values: HIGH, Predicted values: HIGH
Expected values: LOW, Predicted values: LOW
Expected values: MEDIUM, Predicted values: MEDIUM
Expected values: MEDIUM, Predicted values: MEDIUM
Expected values: LOW, Predicted values: LOW
Number correct predicted: 164.0
Number incorrect predicted: 10.0
Percent of correct predicted: 94.25287%
```

Figure 6: The outcome after prediction by the system

D. The scatter plot of violent crimes

This helps to analyze the distribution of violent crimes of given states. It is clearly shown in fig.5 below that some states have minimum violent crimes while others the reverse is true. The more the scatter plots on the state means more violent crimes in that state and the less the scatter plots indicate less violent crimes.

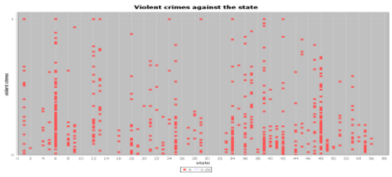

Fig.7: The scatter plot of violent crimes per state

E. Tree visualization

Fig.8 is the graphical representation of the classification tree. A primary goal of data visualization is to communicate information clearly and efficiently. Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable.

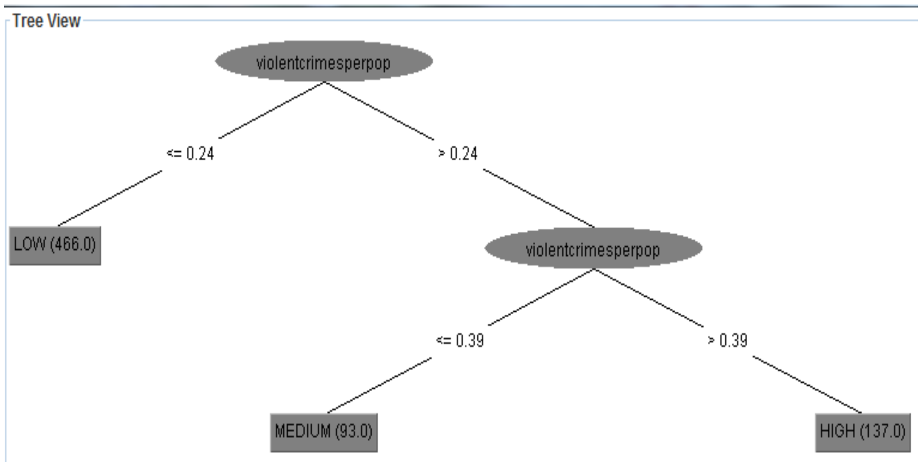

Figure 8: Model classification tree

V. CONCLUSION

A number of classification models were considered as specified in the literature review and compared in analysis stage out of which we chose to use the decision tree (J48) classifier model because of its performance in adapting it to the data collected. We developed a J48 classifier using Waikato Environment for Knowledge analysis (WEKA) Tool Kit and trained it on a preprocessed crime dataset. From the experimental results, J48 algorithm predicted the unknown category of crime data to the accuracy of 94.25287% which is fair enough for the system to be relied on for prediction of crimes and it also takes relatively little time to execute in comparison to other classification algorithms.

ACKNOWLEDGMENT

The authors would like to take this opportunity to show our appreciation to all our colleagues for their support throughout the period of this research. Without such help, this work would not be completed.

REFERENCES

[1] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi and A. Pentland, 'Once Upon a Crime, Towards Crime Prediction from Demographics and Mobile Data', CoRR, vol. 14092983, 2014.

[2] R. Arulanandam, B. Savarimuthu and M. Purvis, 'Extracting Crime Information from Online Newspaper Articles', in Proceedings of the Second Australasian Web Conference - Volume 155, Auckland, New Zealand, 2014, pp. 31-38.

[3] S. O. Adeola, S. O. Falaki and O. Olabode. E-neighborhood Management Architecture for Crime Detection and Control in Nigeria; Science and Technology, 4(2): 17-21 DOI: 10.5923/j.scit.20140402.02. 2014.

[4] Malathi A., Santhosh B.S,. Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters; Global Journal of Computer Science and Technology; Volume 11 Issue 11 Version 1.0 July 2011.

[5] Keyvanpour, M.R., Javideh, M. and Ebrahimi, M.R., Detecting and investigating crime by means of data mining: a general crime matching framework, Procedia Computer Science, World Conference on Information Technology, Elsvier B.V., Vol. 3, Pp. 872-830, 2010.

[6] Nath, S., Crime data mining, Advances and innovations in systems, K. Elleithy (ed.), Computing Sciences and Software Engineering, Pp. 405-409, 2007.

[7] L. P. Walter, M. Brian, C. P. Carter, C. S. Susan and S. H. John. Predictive Policing; The Role Of Crime Forecasting In Law Enforcement Operations; ISBN: 978-0-8330-8148-3. 2013.

[8] Quinlan, J. R., Induction of Decision Trees. Machine Learning 1: 81-106, Kluwer Academic Publishers, 1986.

[9] G. Cybenko. Approximation by superpositions of a sigmoidal function Mathematics of Control, Signals, and Systems, 2(4), 303–314. 1989.

[10] A. S. Rohit, Comparative Analysis of Classification Algorithms on Different Datasets using WEKA. International Journal of Computer Applications (0975 – 8887) Volume 54– No.13. 2012.

[11] Rennie, J.; Shih, L.; Teevan, J.; Karger, D. Tackling the poor assumptions of Naive Bayes classifiers. *ICML*, 2003.

[12] Hand, D. J.; Yu, K. (2001). "Idiot's Bayes — not so stupid after all?". International Statistical Review. 69 (3): 385–399. doi:10.2307/1403452. ISSN 0306-7734.

[13] Narasimha Murty, M.; Susheela Devi, V, Pattern Recognition: An Algorithmic Approach. ISBN 0857294946*, 2011.*

[14] Rennie, J.; Shih, L.; Teevan, J.; and Karger, D., Tackling the poor assumptions of Naive Bayes classifiers. *ICML, 2003.*

[15] V. Vapnik. The Nature of Statistical Learning Theory_ Springer_ NY. 1995.

[16] R. Iqbal, A. A. M. Masrah, M. Aida, H. S. P. Payam and K. Nasim. An Experimental Study of Classification Algorithms for Crime Prediction. Indian Journal of Science and Technology. | Print ISSN: 0974-6846 | Online ISSN: 0974-5645. 2013.

[17] Emmanuel A., Elisha O. O., Ruth W., and Ivan N., Aperformance Analysis of Business Intelligence Techniques on Crime Prediction. International Journal of Computer and Information Technology (ISSN: 2279 – 0764). Volume 06– Issue 02, March 2017.

[18] B. Boehm, "Spiral Development: Experience, Principles and Refinements, Proc. Software Engineering Institute Spiral Development Workshop, p.49, 2000.

[19] J. M. Ngemu, O. O. Elisha, O. O. William, and M. Bernard, M. Student Retention Prediction in Higher Learning Institutions: The Machakos University College Case. International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 04 – Issue 02, 2015.

[20] UCI MachineLearning Repository, Available from: http://archive.ics.uci.edu/ml/datasets.html, 2017.

Turban, E., R. Sharda and D. Delen., Decision Support Systems and Intelligent Systems. 9th Edn., Prentice Hall International, New York, pp: 2, 2011.