IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

The Use of a Decision Tree Algorithm for Classification Purposes in Foretelling the Occurrence of Crime

[1]Dr. Suresh Kopparthi
[1]Professor, Principal
Department of Computer Science and Engineering
Bhimavaram Institute of Engineering and Technology, Bhimavaram.

Abstract—

The crime rate has skyrocketed in the preceding several years. The occurrence of crime is a widespread societal issue that has a negative impact on both living conditions and economic development. As crime rates rise, police departments have a growing need for cutting- edge technology and fresh strategies to enhance crime analytics and strengthen public safety. The use of a decision tree (J48) in the context of law enforcement and intelligence analysis shows promise in mitigating this issue. Data mining is an AI-based technique for gaining insight from big data sets by uncovering previously unknown connections between variables. Decision tree (J48) is one such AI technique. Machine learning is a significant area of study because of its many potential applications. It's no secret that criminology is a prime area for data mining applications. In order to better understand crime, criminologists use a systematic approach called criminology. According to the reviewed literature, the decision tree (J48) algorithm is the most effective machine learning algorithm for prediction of crime data, hence it was chosen for the construction of the study's prototype model of crime prediction. According to the findings of the experiments, the J48 algorithm was able to forecast the unknown category of crime data with an accuracy of 94.25287%, which is good enough for the system to be depended on for the prediction of future crimes.

Keywords: artificial intelligence; classification algorithms; decision tree; J48; crime prediction.

INTRODUCTION

Crime is a widespread societal issue that has a negative impact on both individual well-being and national prosperity [1]. One of the most important factors in deciding whether or not to relocate to a new city and whether or not to visit certain areas [2]. Fear among the populace damages the sense of community, social connections are broken when people avoid specific areas out of habit, people stop venturing out at night, and the town's reputation suffers as a result of crime. People may avoid visiting or even relocate from a neighbourhood if they believe it has a high crime rate. The economy suffers as a result. There are both concrete costs, such as increased demand for police, courts, and correctional facilities, and intangible costs, such as the emotional toll taken on crime victims and the decline in their standard of living. Increasing crime rates are a major issue in many nations nowadays. Actually, researchers are analysing criminals and their actions to learn more about crime and its

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

causes. Since the volume of crime data is growing at an exponential rate, it might provide serious storage and processing challenges. Choosing reliable methods for data analysis is especially challenging because of the inconsistent and inadequate nature of such information. Researchers are prompted to study this data set because of the need to improve crime data analysis. The exponential growth of crime data makes it difficult to store and analyse, among other issues. As a result of the data's inherent inconsistency and insufficiency, questions emerge about the best way to choose appropriate methods for data analysis. To better understand and analyse crime, scientists are motivated to study these sorts of data [3]. The purpose of this study is to use an appropriate machine learning algorithm on crime data in order to forecast whether or not a county would have a low, medium, or high rate of violent crimes.

REVIEW OF THE LITERATURE

Analysis of Crime and Criminal Behaviour The goal of criminology, the scientific study of crime, criminal behaviour, and law enforcement, is to classify different types of criminal activity [4]. It's a major area where data mining may have a significant impact. Researching and identifying criminal behaviour and the circumstances surrounding a crime is what crime analysis in the field of criminology entails. Data mining methods find a natural home in criminology because to the large number of available crime datasets and the inherent complexity of the interactions between them. The initial step in generating additional analysis is identifying criminal characteristics. The insights gleaned from data mining methods are a powerful resource that may aid and assist law enforcement [5]. Data mining is a technology that may aid Law Enforcement Agencies with crime detection difficulties, and according to [6], solving crimes is a hard process that needs human knowledge and expertise. The goal is to use data mining to implement human expertise gained over many years into computational models.

The Predictability of Crime and Why It Occurs

Since criminals often stick to what they know best, there is a lot of data to back up the idea that criminal behaviour can be predicted (statistically speaking) [7]. In other words, they repeatedly engage in the same offences at about the same time and place because they know they have a high chance of success. This is not always the case, but it happens often enough that these techniques are usable in most cases. Routine activity theory, rational choice theory, and crime pattern theory are all prominent explanations for criminal actions. A mixed theory is the result of fusing these various approaches.

Analysing Classification Methods

Algorithms for classifying data that are widely used in making predictions based on past information. An example of a supervised class prediction method is classification. Given a large enough sample size, this method can accurately predict a class's label. Support vector machines, k Nearest Neighbours, weighted voting, and Artificial Neural Networks are just few of the categorization techniques that may be used. All of these methods may be used on a dataset to unearth a collection of models that can be used to predict an unknown class label. As part of classification, data is split into two parts: the training set (also called the dependent set) and the test set (independent set). The machine learning method is first applied to the

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

training set, and then the predictive model is used with the test set. Classification algorithms useful for predicting criminal behaviour are listed below.

Classifier based on a Decision Tree (DT)

In decision tree learning, a tree is used as a prediction model, with each branch representing a set of observations that may be used to determine an item's target value (represented in the leaves). In statistics, data mining, and machine learning, it is one of the predictive modelling methodologies employed. Classification trees are a kind of tree model in which the leaves represent class labels and the branches indicate conjunctions of characteristics that lead to those labels, and in which the target variable may take only a limited range of values. Regression trees are a kind of decision tree where the dependent variable may take on a continuous range of values, most often represented by real numbers. A decision tree is a useful tool for representing choices and decision- making processes graphically and clearly, which is useful for decision analysis. A decision tree is a data mining tool used to visually represent and analyse complex sets of information (but the resulting classification tree can be an input for decision making). The C4.5 [8] method, an expansion of Quinlan's previous ID3 algorithm, is used to create a decision tree. Nodes in the tree are chosen using an entropy metric. Qualities were chosen in descending entropy order due to the fact that more uncertain outcomes are associated with greater entropy attributes. In order to "learn" a tree, it is necessary to partition the source data into smaller subsets, each of which is tested for its value in a single attribute. Recursive partitioning is the process of applying this step to all of the resulting subsets. Spaces that have been partitioned and those that have not utilising recursive partitioning (also known as recursive binary splitting) are shown as examples in the picture. When the subset at a node has the same value of the target variable, or when further splitting does not improve the predictions, the recursion is complete. By far the most popular approach to learning decision trees from data, top-down induction of decision trees (TDIDT) is an example of a greedy algorithm [9]. Decision trees are used in data mining to help in the description, classification, and generalisation of data sets via the use of a variety of mathematical and computational approaches. Records are the medium for storing data.

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

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, x1, x2, x3 etc., that are used for that task.

ISSN PRINT 2319-1775 Online 2320-7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

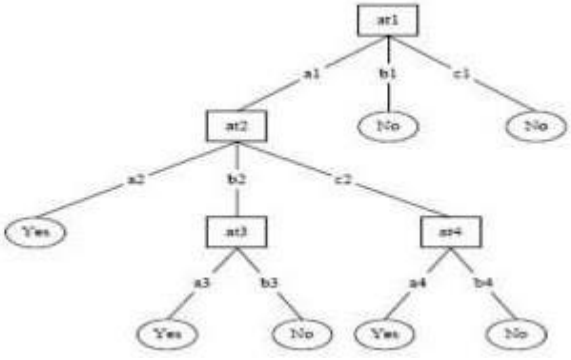


Figure 1: Decision tree, Source: Gama et al, 2003

Assembling Perceptron Layer by Layer (MLP)

Multilayer Perceptron's are a kind of feedforward artificial neural network model that are used to convert data sets into meaningful outputs. To train its networks, MLP uses the supervised learning method of back propagation.

As it has three or more layers of nonlinearly- activating nodes (input, output, and hidden layers), the Multilayer Perceptron belongs to the category of "deep neural networks" [10]. As a whole, the layers are composed of smaller parts. The network's inputs are the properties that were gathered for each tuple in the training set. The units that make up the input layer receive the inputs all at once. These inputs are routed via the input layer before being sent into a hidden layer, which is a second layer of "neuron like" units that receives them weighted and in parallel. A new hidden layer may be fed the outputs of the previous one, and so on. However, in reality, just one hidden layer is employed, and this number is completely arbitrary. Reducing the number of input units in a neural network and speeding up the training period are both ways to boost classification accuracy [11]. A multi-layer neural network is made up of many individual units (neurons) connected to one another in a certain way. In order to be classified as a deep neural network, the multilayer Perceptron requires three or more layers (input and output layers plus one or more hidden layers). Each node in one layer has an associated weight with every node in the next layer since an MLP is a Fully Connected Network. Whether should be understood as the weight from I to j, or vice versa, is a point of contention, as does whether the input layer should be counted as one of the layers or not. An input layer, one or more hidden layers, and an output layer make up a multilayer feed-forward neural network. Fig. 2 depicts an example of a multilayer feed-forward network.

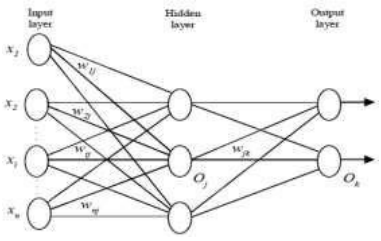


Figure 2: A multilayer feed-forward network. Source: Rohit, 2012 According to [12], each layer is made up of units.

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

The network's inputs are the properties that were gathered for each tuple in the training set. The units that make up the input layer receive the inputs all at once. These inputs are routed via the input layer before being sent into a hidden layer, which is a second layer of "neuron like" units that receives them weighted and in parallel. A new hidden layer may be fed the outputs of the previous one, and so on. However, in reality, just one hidden layer is employed, and this number is completely arbitrary. Back propagation is, at its heart, just a fast and accurate way to compute all the derivatives of a single goal variable (such pattern classification error) with regard to a vast number of input values (such as the parameters or weights in a classification rule). We can increase classification accuracy by decreasing training time for neural networks and simplifying their input unit architecture.

Classifiers using naive Bayes

There is a family of straightforward probabilistic classifiers known as Naïve Bayes classifiers, which use Bayes' theorem under the naive assumption of complete feature independence. The 1950s marked the beginning of a period of intensive research on Naive Bayes. It was first introduced to the text retrieval community in the early 1960s under a different name, and it continues to be 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.) based on word frequencies as the features. In this area, it may compete with more sophisticated algorithms like as support vector machines [13] given the right amount of pre-processing. Automatic medical diagnosis is another field where this helps out. The number of parameters needed for a Naive Bayes classifier is proportional to the number of variables (features/predictors) in the learning problem, making it extremely scalable. Unlike the costly iterative approximation employed by many other kinds of classifiers, maximum likelihood training may be done simply evaluating a closed-form expression, which requires linear time. Some other names for naive Bayes models include simple Bayes and independence Bayes. All of these terms allude to the fact that the classifier makes use of Bayes' theorem in its decision process, yet naive Bayes is not (necessarily) a Bayesian approach [14].

It's all thanks to SVMs (Support Vector Machines)

For the purposes of classification and regression analysis, support vector machines (SVMs; sometimes known as support vector networks) are supervised learning models with accompanying learning algorithms. An SVM training method is a non-probabilistic binary linear classifier; given a collection of training examples, each of which is labelled as belonging to one of two categories, it constructs a model that assigns future instances to one of the two categories. The examples in an SVM model are points in space that are mapped such that there is a clear, as large as feasible, difference between the instances in the various categories. Then, fresh instances are projected into this region, and their classification is predicted according to which side of the chasm they land on. By the mapping its inputs into high dimensional feature spaces, SVMs may easily execute a non-linear classification in addition to linear classification. When data are unlabelled, supervised learning cannot be used; instead, an unsupervised learning strategy is necessary, which seeks to discover natural grouping of the data to groups and then maps additional data to these created groups. When data are not labelled or just partial data are labelled, support vector clustering, a clustering technique that improves upon support vector machines, is widely employed in industrial applications as a pre- processing step before a classification run [15].

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

Examining the efficacy of categorization systems for predicting criminal behaviour

The accuracy of the decision tree method (83.9519%) and the Nave Bayes algorithm (70.8124%) used to analyse crime data in [16] was, respectively, 83.9519% and 70.8124%. Therefore, he reasoned, decision trees are superior than Naive Bayes. [17] The accuracy of the decision tree (J48), Naive Bayes, Multilayer Perceptron, and support vector machine was 100 percent, 89.9425%, 100%, and 93.6782%, respectively, with execution times of 0.062 seconds, 0.014 seconds, 9.26 seconds, and 0.66 seconds, when tested on crime data. As a result, decision trees were faster to execute and more efficient overall. To this end, the suggested BI system would use the researcher's preferred method, a decision tree algorithm (J48), due to its fast and precise performance.

Theoretical Framework for the Suggested System

This research determines the best method based on an analysis of the aforementioned literature and prior research in the topic. Below is a diagram depicting the conceptual model constructed using these methods for this study.

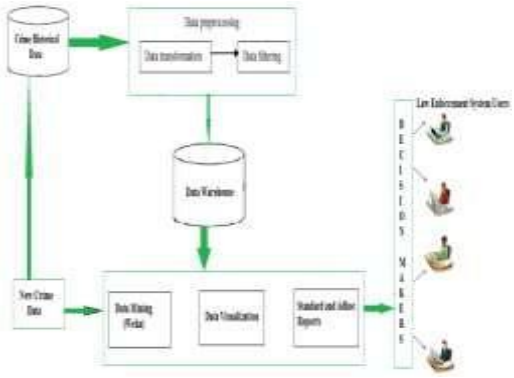


Figure 3: Block diagram of the proposed crime predictive System

CONCEPTUAL FRAMEWORK AND RESEARCH APPROACH

Every step of the process, from initial requirements gathering to final code deployment, was guided by the spiral model.

Summary of the Spiral Model

The Spiral Model is a systems development lifecycle model that includes a thorough procedure for describing, creating, and implementing prototypes [18]. It is a hybrid of the Prototyping Model and the Waterfall Model. Projects that fall under those categories tend to choose the spiral model [19].

How Machines Do It

Since labelled training data was readily accessible, we investigated a machine learning model that relied on supervised learning (classification) approaches. The aim of classification is to assign a new observation to one of many predefined groups (subpopulations) based on its similarities to a predefined collection of examples (the training set) for which the group membership is already known. Crime prediction also made use of a

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

decision tree classifier. Our procedure begins with gathering raw data, then moves on to pre-processing, model construction using training data, and finally model assessment with test data. The model was trained and validated, and then it was utilised to evaluate new data.

Where We Get Our Information

The data utilised in this research is a genuine and original one. We downloaded the dataset from the UCI machine learning repository. Crime and Communities is the name of the data collection.

There are 128 characteristics and 1994 cases overall in this collection. The data in this collection is numerical and has been standardised. The UCI machine learning repository website [20] contains all 128 characteristics and their descriptions in full.

Attribute choice

My analysis's goal (crime forecasting) did not call for the use of all the recorded variables; so, I had to prepare, reduce, and pre-process the data. While performing data reduction, it is important to keep in mind that you shouldn't remove any characteristics from the dataset that are crucial to the classification process. Eliminating the unnecessary factors was a must. Only 12 of the original 128 qualities were really used in the end. Attribute and feature selection may be performed in a number of ways, but the most common approach relies on a human's intuitive grasp of the data. It is OK to rely on human understanding when making judgements about a large number of characteristics, so long as care is made to choose just those attributes that do not have any missing values.

Contextual factors

Country, residents, and numbers Median Income (Med Income), Median Family Income (Mediatic) and Per Capita Income (Percipience) (Per capita income), Underpot (Number of persons beneath the poverty level) (Number of people under the poverty level), The percentage of adults aged 25 and up who did not complete high school, the percentage of adults aged 25 and up who did not complete high school, the percentage of adults aged 25 and up who did not complete high school, the percentage of adults aged 25 and up who did not complete high school, the percentage of adults aged 25 and up who did not complete high school, the percentage of adults aged 25 and up who did not complete high school, the percentage of adults aged 16 and over who are unemployed, the percentage of adults aged 16 and up who are (Crime categorization in to three categories, namely). The new nominal characteristic may take on one of three possible values: Low, Medium, or High. Crime is considered "Low" if the proportion of violent crimes committed in a given population is less than 25%; "Medium" if the proportion is between 25% and 40%; and "High" if the proportion is either more than 40% or equal to it.

Methods and Software for Doing Modelling

In this analysis, we focused on a model that used a supervised learning (classification) approach. WEKA, a knowledge-analysis programme available online for free download and usage under the GNU licence, was the programme of choice. Multiple machine learning algorithms are supported by WEKA. JAVA was used for both the display of findings and the creation of the prototype, and Javad will be used to store the data. This Java-based

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

release (Weka 3.8.0) sees usage in a wide variety of contexts, including pedagogy and science. Weka is capable of doing many of the common data mining operations, including cleaning, grouping, analysing, visualising, and selecting features. Weka's premise for its methods is that data may be provided in the form of a single flat file or relation, with a defined set of characteristics for each data piece (normally, numeric or nominal attributes, but some other attribute types are also supported).

A REVIEW OF THE FINDINGS AND AN EXPLANATION OF THEIR SIGNIFIC

A software prototype was developed for this study's experimental purposes, and its design and implementation are described here. The data set utilised, the programming methodologies opted for, and the testing procedures for the system are described in detail.

Assembled Data Used for Instructional Purposes

As shown in Fig.2, we utilised a data set with known output values as training data to create the model. In contrast, this model splits a whole training set into two halves, with the first portion serving as the basis for the model's creation (the training set), and the second part serving as an instant accuracy check (the test data set).

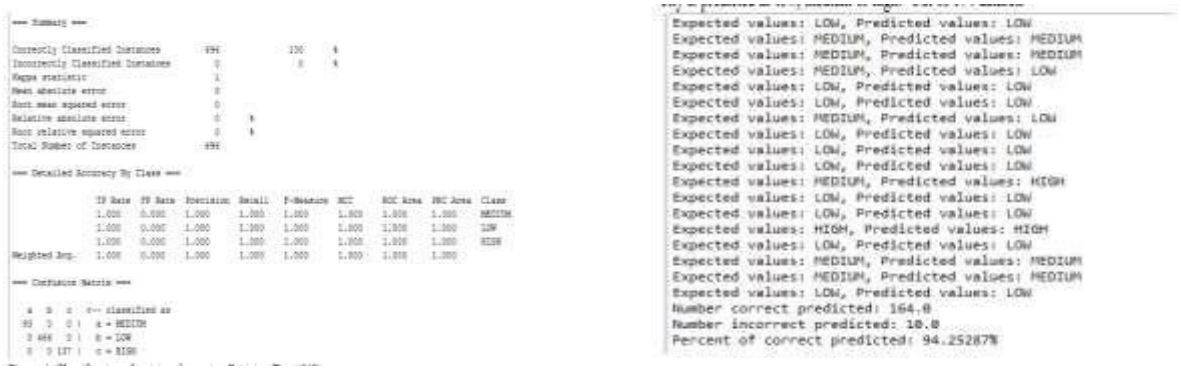
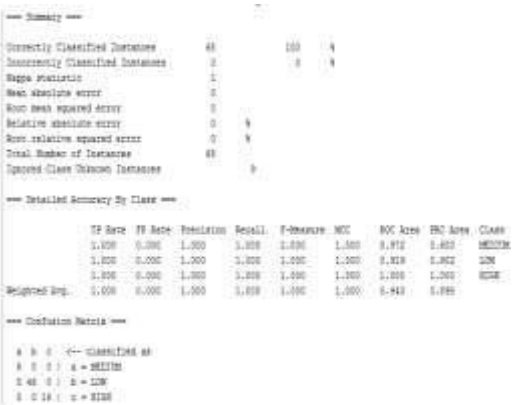


Figure 4: Classification of training data using Decision Tree (J48)

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.



ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

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

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.

The scatter plot of violent crimes

This helps to analyse 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.

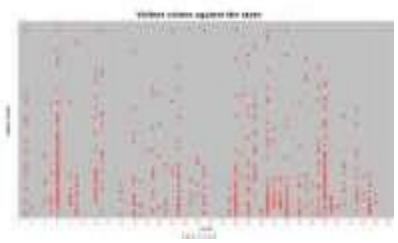


Fig.6: The scatter plot of violent crimes per state

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 analyse and reason about data and evidence. It makes complex data more accessible, understandable and usable.

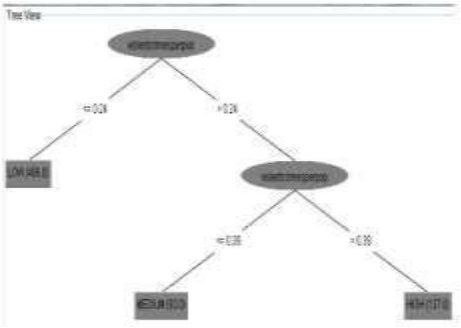


Figure 7: Model classification tree

CONCLUSION

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

In the analysis phase, we took into account the various classification models outlined in the literature review and compared them to one another before settling on the decision tree (J48) classifier model due to its superior performance when applied to the available data. Using the Waikato Environment for Knowledge analysis (WEKA) Tool Kit, we created a J48 classifier and trained it on a cleaned-up crime data set. Experimental findings showed that the J48 algorithm accurately identified the unknown category of crime data with an accuracy of 94.25287%, making it a viable candidate for use in crime prediction systems. It also runs quickly in contrast to other classification algorithms.

REFERENCES

[1] A. Bogomil, 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

ISSN PRINT 2319-1775 Online 2320-7876

 UGC CARE Listed (Group -I) Journal Volume 10, Iss 3, Mar 2021

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

[17] 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.

[18] 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.

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

[20] 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.