Original Article

Identity Investigation System with Suspect Predication

*Appuhamy H.A.R. [#1], De Silva R. [*2], Dhammearatchi D. [*3],*

#Department of Information Technology, CINEC Campus
Millennium Drive, IT Park, Malabe, Sri Lanka
[1]ann99roshanieappuhamy@gmail.com

Abstract

Crime is one of the major problems encountered in any society. Thus, there is an important need for security agents and agencies to battle and eradicate crime. Sri Lankan police is responsible for investigating and detecting criminals of any crime committed within the country. Crime is an act usually deemed socially harmful, specifically defined, prohibited and punishable under criminal law. Crime rates are rapidly increasing and changing. Crime prediction is very important in any country. In Sri Lanka, crime detection among people is very crucial to Department of police office. The purpose of Identity investigation system with suspect prediction that tracks the investigation status of criminal cases with logs and also predicts primary suspects. This project is mainly based on Android Studio and WEKA tool. This project has main two parts. In system has suspect prediction part. In this system it uses data mining techniques for prediction process. For that purpose, mainly system uses WEKA Tool. Identity investigation process is second main part of this project. It is doing by using SQL functions. It uses Short Message Service (SMS) and other functions for data security. Basically, mobile application development doing by Android Studio.

Keywords: SMS, Alarm, Android Application, Data Mining, WEKA Tool, SQL Function

Introduction

At the present crimes are increasing highly in Sri Lanka. In Sri Lanka, police department is the largest unit for preventing crimes, maintaining law orders, rules and peace throughout the country. However, problem with the Sri Lankan police is that they are still using the traditional manual process to keep and analyse the records of crime and criminals. On the other hand, criminals use more advance technologies to commit the crime in more tactful ways. Crime rates in Sri Lanka are increasing enormously among countries and their variations in this dimension is orders of magnitude greater their variation through the time in a given country. [1]. Over the previous few years most of the countries of the world has experienced a remarkable increase in rate. There is no explicit reason for any disorder for criminal activities. Generally, society, cultural factors, totally different family systems, political influences and enforcement are responsible for criminal activities of an individual. Crime rate is growing in Sri Lanka. Crime is also found in numerous forms. Organized crime includes drug trafficking, shooting, concealing, extortion, and murder for rent, fraud, human trafficking. [2] Several criminal activities result in political violence, religiously impelled violence, terrorism, and abduction. The opposite sort of crimes includes homicide, robbery, assault etc. Corruption may

be an important drawback that is faced by Sri Lanka. Thus, there is a large need for

situation. Sri Lankan police is responsible for investigating and detecting criminals of any crime committed within the country. This research aims to develop a smart mobile system for criminal information analysis for Identity Investigation System with Suspect Predication. The main purpose of the system is to correctly classify the crime data and use the most appropriate data mining algorithm to reveal the hidden knowledge from crime data. This knowledge can be very useful in identifying possible suspects of the crime. Selection of appropriate data mining algorithm will be done by J48 decision tree algorithm obtained from WEKA mining software.

Research Methodology

In order to answer the research questions, the author has followed this process.

A. Requirements Gathering

Requirement management is the most important area of this research project. In this Process author has identified the major requirements of this project.

- Create a dataset for the scenario.
- Identify the suitable algorithms for scenario.
- Identify the suitable prototype for scenario.

To address first requirement of the project the author has to generate **suitable dataset for process.** Data Mining (DM) is the process for automatic discovery of high level knowledge by obtaining information from real world, large and complex data sets [3], and is the core step of a broader process, called Knowledge Discovery from Databases (KDD). In addition to the DM step, the KDD process includes application of several pre-processing methods aimed at facilizing application of DM algorithms and post processing methods for refining and improving the discovered

developing smart solution for this knowledge. This idea of automatically discovering knowledge from databases present a very attractive and challenging task, both for academia and industry.

First of all, author generate the crime dataset through two sequential steps

(1) Data Extraction extracts the unstructured crime data from various crime Web sources, namely—. National Crime Records Bureau (NCRB), Committee to Protect Journalists (CPJ) and other Web sources during the period of 2000–2012[4].

(2) Data set cleans, integrates and reduces the extracted crime data into structured 10,000 crime instances (.csv format). The structured crime dataset is represented using 11 crime attributes. The formulated dataset is implemented using two JAVA tools— (1) Netbeans for crime detection, criminal identification and prediction. (2) WEKA for crime verification. Some of the crime attributes are mentioned in following figure.

Figure 1:Train Data Set
(Source: Created by author)

To address second requirement of the project the author has selected **Decision Tree Analysis on J48 Algorithm.** [5] Basically, decision makers prefer a decision tree because it is not complex as well as easy to understand. Tree complexity has its effect on its accuracy. Usually, the tree complexity can be measured by a metrics that contains: the total number of nodes, total number of leaves, depth of tree and number of attributes used in tree construction.

[6] Decision trees are the most powerful mining. It includes the technology of research large and complex bulk of data in order to discover useful patterns. This idea is very important because it enables modelling and knowledge extraction from the bulk of data available.
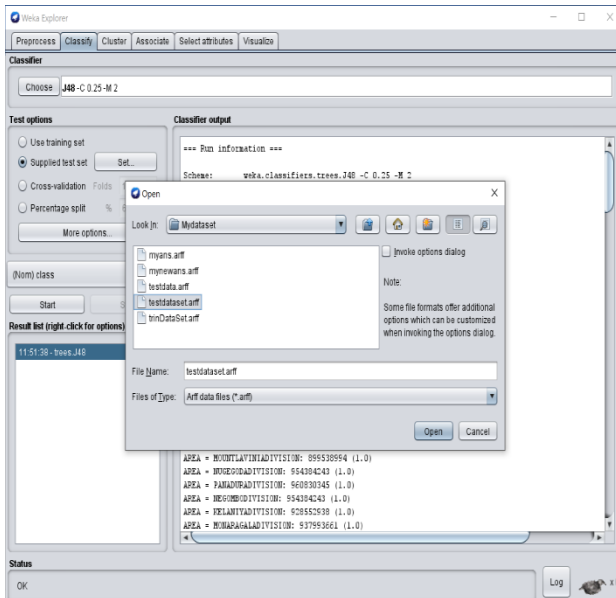

Figure 2: Select J48 Model
(Source: Created by author)

To address third requirement of the project, the author has selected the good **proto type** for the project.

To implement this research the author has selected the **prototyping** model. The prototyping model is a systems development method in which a prototype is built, tested and then reworked as necessary until an acceptable outcome is achieved from which the complete system or product can be developed [7].

B. Analysis

In this analysis part author carries deep investigation about the algorithms. To develop a accuracy high system, author has to select good algorithm for system. Basically, the author has selected two major algorithms and comparison has done by considering the accuracy levels.

approaches in knowledge discovery and data Classification technique is used to find the common properties among different attributes of crime and criminal dataset and organizes them into different predefined classes. Classification is a data mining technique which is often used for predicting crime patterns and it also can reduce the time required to identify these crime patterns. However, classification technique requires proper training and testing data because a certain number of missing values may limit the prediction accuracy. [8] For example, we can apply classification in crime application that given all past records of criminals who commit the crime, predict which current criminals are probably to commit the crime in the future.

Firstly, author should have a set of training instances. This uses a J48 decision tree algorithm and merit selection criteria to choose the best splitting attribute to create a branch. Thus, author has two partitions. Algorithm will apply same top-down analysis to make further more partitions. One of the stopping criteria is when all the attribute's values belong to a single class. There is only difference in splitting criteria if do comparison between multi and univariate tree construction. [9]. 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 crime prediction system.

In second section, the author has tried to train data set with k-means clustering, the results of k-means clusters are verified with WEKA. WEKA verifies an accuracy of 52.62 and 52.99 % in the formation of two crime clusters using the selected crime attributes.

Error Calculation:

Cluster 1 = MOD(2,445 - 2,289) / 2445 = 0.0638

Cluster 2 = MOD(2,593 - 2,749) / 2593 = 0.0601

So, Accuracy Measure: ..

Cluster 1 = 52.62 %

Cluster 2 = 52.99 %

It is obvious that, the best algorithm is **J48** and it is fair enough for the system to be relied on for crime prediction system.

C. Identifying Business Values

The system is designed to aid investigation teams to work collectively on cases, coordinate and also speed up the process by suggesting logical suspects based on data provided. It is their sole responsibility to enforce the law, find and apprehend irresponsible society, reduce and curtail any and every form of indiscipline. Some of the crimes in this situation include: burglary, sexual harassment (and/or rape), abuse of drugs, alcoholism, homosexuality, misuse and abuse of school properties, disobedience of school rules and regulations, stealing and many other crimes included in the Sri Lanka Conduct/Rules and Regulations [10]. After a crime is committed, it becomes the duty of the assigned police officer to forecast the potential suspects of that crime, perform a series of investigations, apprehend and then prosecute the real criminal. Implemented system helps to overcome this situation.

D. Design

Figure 01 and 02 explains the design of the system through a use case and the data flow diagram architecture.

These are the main functions of this suspect prediction system.
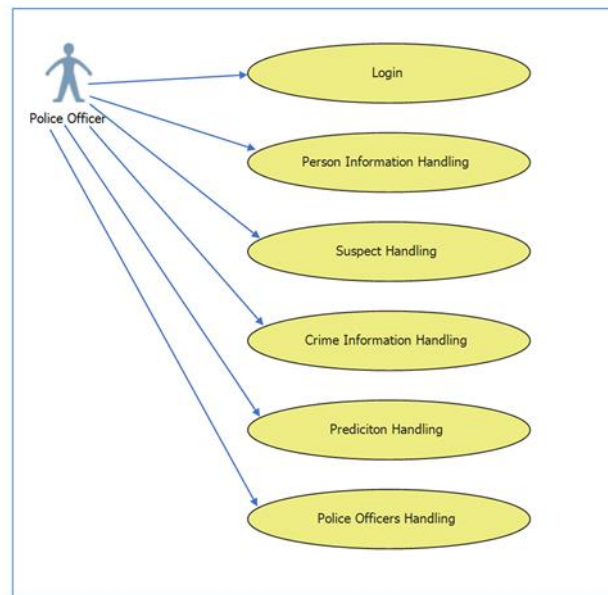
Figure 3: Usecase Diagram for system.

(Source: Created by author)
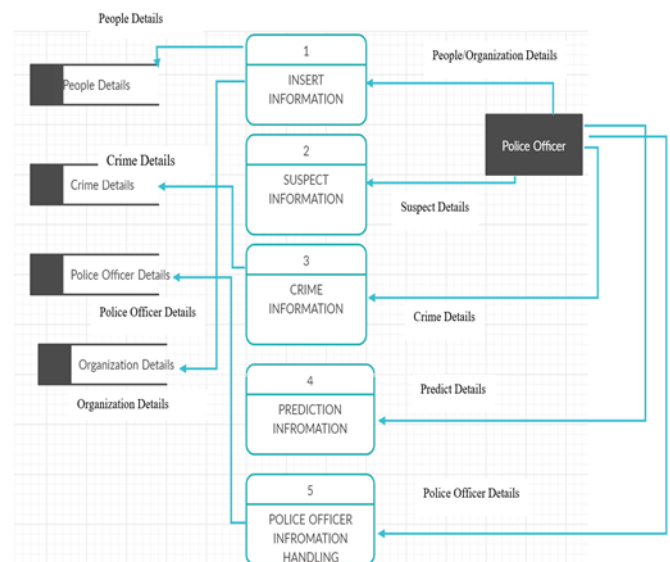
Figure 4: Data Flow Diagram.

(Source: Created by author)

UML use case is a popular modelling technique especially in object-oriented based software development. A UML model produced consists

of different diagrams, views of different level of abstraction and contributes from different stakeholders and modellers. These open to consistency problems between diagrams [11]. In this system the main actor is the Police office. Police officer. Police officer can insert information, prediction process to the system.

Data Flow diagram basically shows all data stores and basic operations of the system. Mainly police officer can access and handle all data stores. Above diagram shows all data flows of this system.

E. Implementation

In this project author has used main three components. Data Mining (WEKA Tool), database, Android application.

In implementation stage, basically authors have created this project according to the following process flow chart. In a software implementation, it is often easier to trade off reliability for performance because the software can be changed or adopted to specific needs. Mainly this system develops under WEKA Tool. It has been indicated through Figure 03.
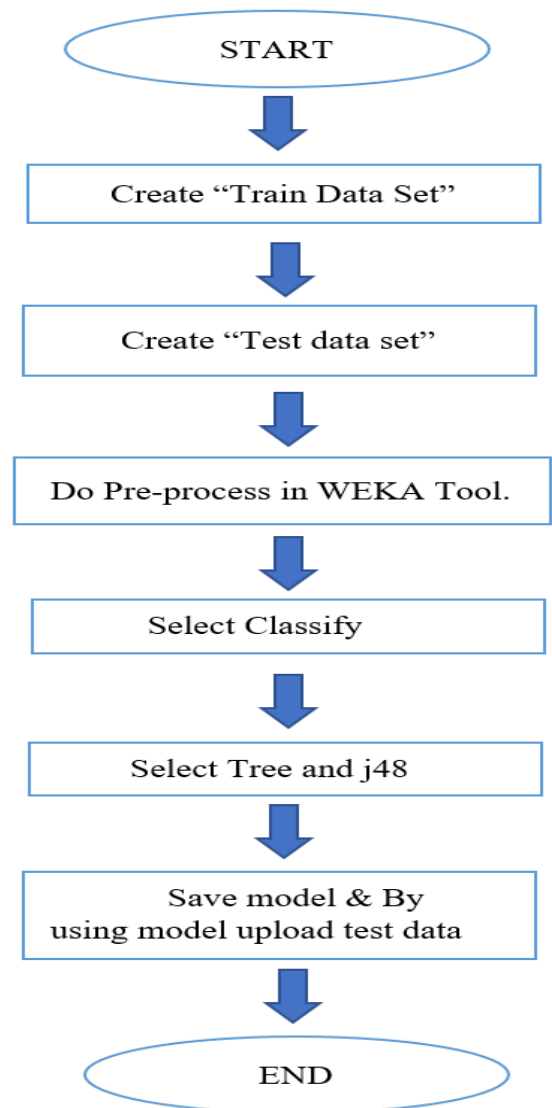
Figure 5: Process Flow chart Diagram.

(Source: Created by authors)

STEP 01- Preprocess
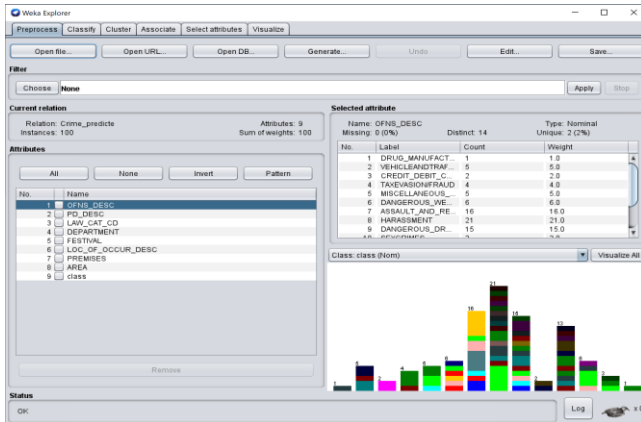
Figure 6-Preprocess
(Source: Created by authors)

STEP 02- In Classify · Select tree · j48
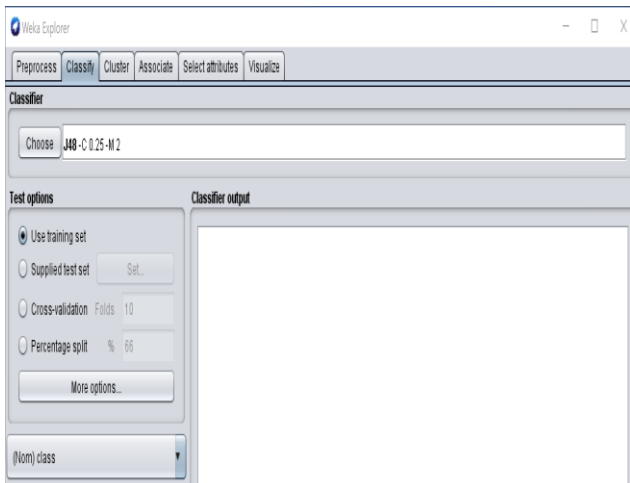
Figure 7-Classify
(Source: Created by authors)

STEP 03- Save Model
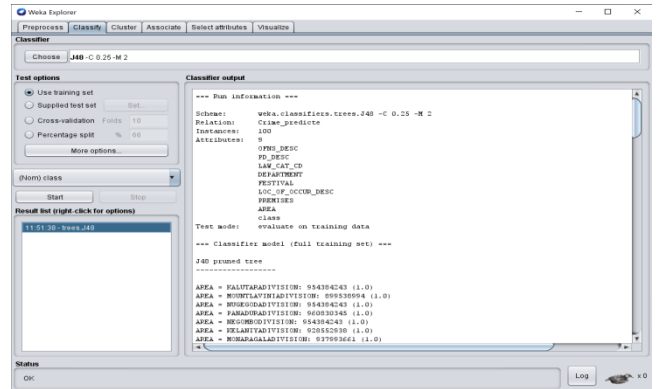
Figure 8-Save Model
(Source: Created by authors)

TEP 04- By using model upload test data file
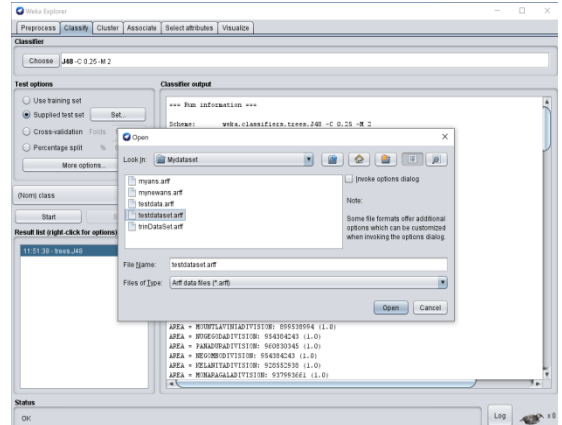
Figure 9-upload test data file (Source: Created by authors)
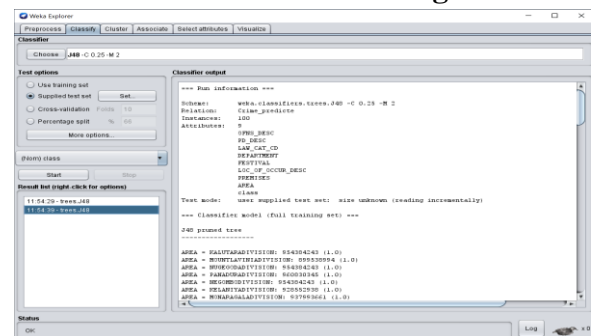
STEP 05- Start Predicting.

Figure 10-Start Predicting (Source: Created by authors)

Figure 11: Admin Page (Source: Created by authors)

F. Testing

In computer programming, unit testing is a software testing method by which individual units of source code are tested to determine whether they are fit for use. A unit is the smallest possible testable software component. Usually, it performs a single cohesive function.

In this project authors mainly did test major two parts.

01) Mobile Application in Android Studio.

Authors did simple testing to my mobile application and all functions are checked under it.

02) Data Mining WEKA TOOL

Results and Discussion

This content covers the results that were achieved from the research project and what were the new approaches found to address further research in the undergraduate context.

This is the final outcome of this product. By using this Admin page, the users can move to prediction page
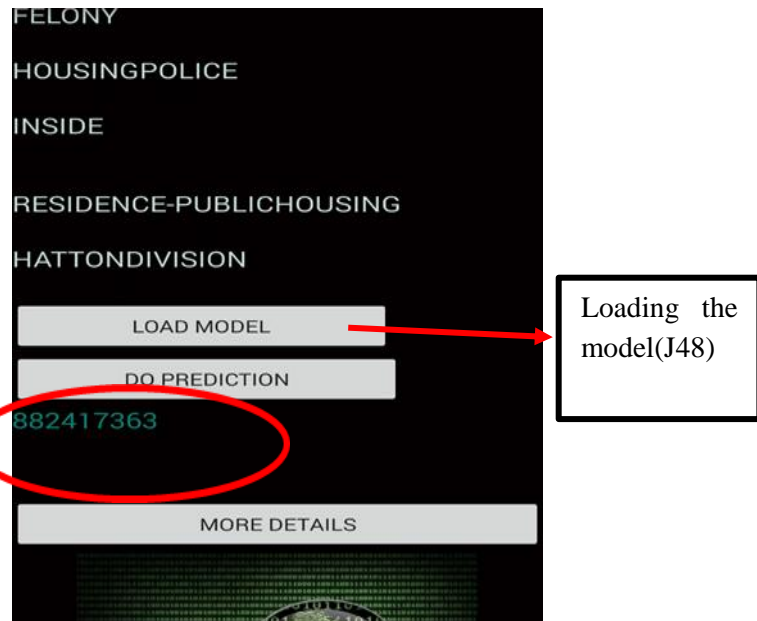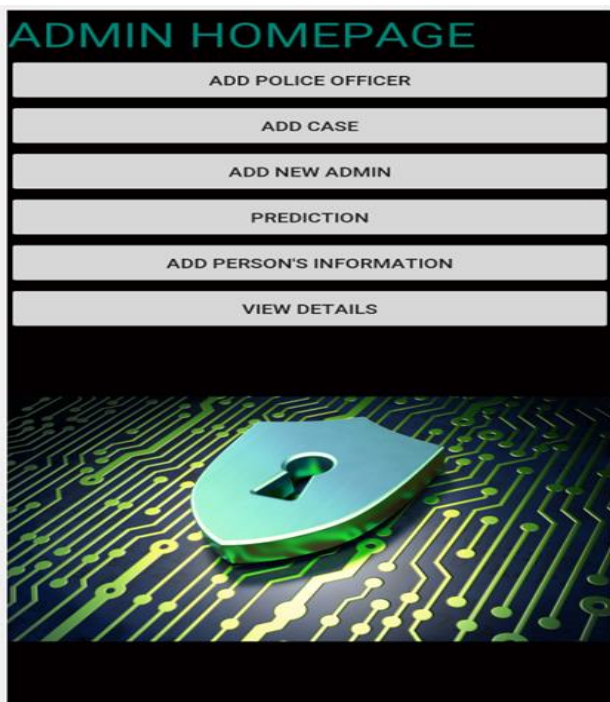

Figure 10: Prediction Page
(Source: Created by authors)

Identity Investigation System with Suspect Predication searches for immanent patterns and relations in the given crime data by using the J48 techniques. In this section, the results of J48 are verified with WEKA. Here the classification result of J48 – a decision tree is shown using parameters such as correctly classified instances, true positive rate, and false positive rate [12]. This technique provides an overview of large amount of the crime data and facilitate in handling, searching and retrieving of the desired crime information. Identity Investigation System with Suspect Predication can also be useful for crime prevention

Future works and Conclusions

Implemented system presented an efficient and affordable method for real time problem solve system this system can be updated to the extraordinary level. Therefore, authors are strongly recommended this system will be solve numerous problems regarding its administration in high level.

In order to use this system research, authors have not enough data to data mining to get good prediction. Though the data for data mining is a limitation. Recommendation to those who willing to develop this system further as follows;

- Add more data and do the data mining to get good result.
- Develop the system to get more accurate result.
- Add international prediction algorithms to this system.
- Add more security features.
- Enhance number of users.
- More user-friendly interfaces.
- Add more algorithms.
- Increase accuracy.

Above changes can add more quality for this system.

As a future extension of this study, a web based criminal identification system is proposed which will give information of criminal at a single click from any place. This system will also serve the purpose of crime information sharing and investigators can access the criminal information from any place.

Declarations

Study Limitations

When developing this project, authors followed sample data mining projects. Authors achieved their target scope and authors face some difficulties when working with data sets. Here for data mining purpose authors needed very large amount of data records and when handling that data set, authors faced some limitations. In data mining process, it is a known fact that data mining collects information about people using some market-based techniques and information technology. To run this data mining-based project authors required high performed machines. These are main identified study limitations of this project.

Acknowledgements

References

[1] S. Ghosh, "Formal approach on modeling and predicting of software system security: Stochastic petri net", Journal of Artificial Intelligence and Data Mining, vol. 3, no. 1, 2015. Available: 10.5829/idosi.jaidm.2015.03.01.08 [Accessed 23 November 2019].

[2]. Becker RF (2008) Criminal investigation, 3rd edn. Jones and Bartlett Learning Publishers, Burlington, MA [Accessed 18 August 2020].

[3]. National Crime Records Bureau, http://ncrb.gov.in. 6 Committee to Protect Journalists, http://www.cpj.org [Accessed 23 November 2019].

[4]. J. S. Aguilar-Ruiz, R. Giráldez, and J. C. Riquelme. (2007). Natural encoding for evolutionary supervised learning. IEEE Transactions on Evolutionary Computation, 11(4):466–479[Accessed 23 August 2020].

[5]. Korting, Thales S. "C4.5 algorithm and Multivariate Decision Trees." 5. [Accessed 13 August 2020].

[6]. Quinlan, J. R. "Improved Use of Continuous Attributes in C4.5." 14. [Accessed 21 August 2020].

[7]. H. Fredrick David and A. Suruliandi, "SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES", ICTACT Journal on Soft Computing, vol. 7, no. 3, pp. 1459-1466, 2017. Available: 10.21917/ijsc.2017.0202 [Accessed 23 August 2020].

[8]. "Criminal Investigation Tracker with Suspect Prediction Project", Nevon Projects, 2020. [Online]. Available: https://nevonprojects.com/criminal-investigation-tracker-with-suspect-prediction/. [Accessed 20 August 2020].

[9]. Korting, Thales S. "C4.5 algorithm and Multivariate Decision Trees." [Accessed 20 August 2020].

[10]. "Criminal Investigation Tracker with Suspect Prediction Project", Nevon Projects, 2020. [Online]. Available: https://nevonprojects.com/criminal-investigation-tracker-with-suspect-prediction/. [Accessed: 08 September 2019].

[11]. Sommerville, I. (2004). Software Engineering 7 (Seventh ed) (Second ed.). Harlow: Pearson Addison Wesley. [Accessed 21 August 2020].

[12]. Ian H. Witten, Eibe Frank, "WEKA – Machine Learning Algorithm in Java" [Accessed 21 August 2020].