Long paper

Development of Crime Reporting System to Identify Patterns of Crime in Laguna

Mia V. Villarica
College of Computer Studies, Laguna State Polytechnic University, Philippines
mia.villarica@lspu.edu.ph
(corresponding author)

Francis F. Balahadia
College of Computer Studies, Laguna State Polytechnic University, Philippines
francis.balahadia@lspu.edu.ph

Jonardo R. Asor
College of Computer Studies, Laguna State Polytechnic University, Philippines
jonardo.asor@lspu.edu.ph

Gene Marck B. Catedrilla
College of Computer Studies, Laguna State Polytechnic University, Philippines
genemarck.catedrilla@lspu.edu.ph

ABSTRACT

Purpose – This study developed an online crime reporting system that uses artificial intelligence to analyze crime incident reports to provide up-to-date crime statistics, map crime hot locations, and manage dynamic databases.

Method – The knowledge discovery process in databases (KDD) was utilized for the model development. Scrum, an agile development technique, proved helpful in the iterative and

gradual development of the system. In addition, several ICT technologies were applied, such as geo-mapping and pattern analysis utilizing the data mining technique. The user evaluation tool was composed of Technology Acceptance Model for the criteria and ISO/IEC 25010 software metrics for the sub-criteria.

Results – Based on the patterns generated from the criminal records data set, the researchers used machine learning in a prediction model generated using the Decision Tree algorithm, revealing several important insights about the incidences of non-index crimes in Laguna. The findings suggest that date, time, and location factors are the best predictors of crime occurrences. Moreover, the researchers agree with the respondents' comments and suggestions that the crime map should include a variety of graphical representations such as a table ranking the crime rates from highest to lowest and a pie graph showing the comparable data of analytics crime per town and cities to make the system more interesting to any type of user. This is true for the crime analysis website's crime map for public access user review, which ultimately came out to be acceptable.

Conclusion – The online crime reporting system provides various functions and features for various users. This can be used to raise people's awareness regarding dangerous locations and help agencies predict future crime in a specific location within a particular time.

Recommendations – PNP-Laguna would gain the most from the project, it is recommended that the PNP-Laguna, municipal police stations, and the LSPU work efficiently together to continue expanding the crime analysis website. The website offers analytics for decision-making support in addition to covering crime management information systems across police agencies for decentralization. Continued experimentation on index and non-index crime datasets to develop intelligent systems capable of forecasting must be explored, as the index crime dataset demonstrated a highly good result in the project.

Research Implications – The academe and local law enforcement agencies must collaboratively develop strategies that reinforce the importance of community engagement. The system provides public access through geo-mapping that supports the community, or locally based crime prevention, instead of targeting individuals, targets areas where the risks of becoming involved in crime or being victimized are high.

Keywords – crime incidents, crime report, crime patterns, Laguna, Decision Tree algorithm, KDD

INTRODUCTION

Today, technology has permeated every facet of our culture. It has become a vital part of our lives due to the numerous benefits it provides—education, medicine, transportation, and, most crucially, law enforcement all benefits from technological advancements. Even

a brief inspection of the evolution of our efforts to prevent crime demonstrates that technology – or more precisely, technological innovation – has been the driving force behind the reform of crime prevention and crime control strategies, both by private citizens and concerned groups and by traditional police agencies (Reichert, 2001; Chan, 2001; Harris, 2007).

In recent years, the Philippines has seen an increase in crime. According to the Philippines, the index Crime is the primary delinquent record in the Philippines National Police (PNP) due to unplanned urbanization, migration, a high rate of unemployment, and poverty. As for the PNP's techniques, they take immediate action when a concerned citizen reports and accuses someone of engaging in criminal activity or being suspicious of anything, and they also patrol areas where there is a risk for crime. The lack of preparedness and awareness of public safety is one of the factors in the Government issue. Despite these rising rates, this project places a premium on crime prevention. It could be reducing criminality in a specific area's percentage or preventing risk before it occurs using data mining technology. The research analysis presented in the 14th National Convention on Statistics in 2019 organized by Philippine Statistics Agency (Baculinao & Ceballos, 2019) revealed that Cavite, Laguna, Batangas, Rizal, and Quezon (CALABARZON) should focus on policies that will help prevent cases of crimes and improve crime solution efficiency.

As the new century accelerates technological advancements, real-time reporting of crime situations is now feasible via mobile and web applications. Numerous tools and techniques are used to address crime-related issues but implementing a comprehensive crime information system that manages crime report databases, criminal records retrieval, and the integration of crime prediction models with spatiotemporal patterns of incidents for forecasting is infrequent. Roth et al. (2013) define crime analysis as the systematic collection, preparation, evaluation, and dissemination of data about criminal activities to aid law enforcement in their goal. It is a collection of systematic and analytical processes for determining crime patterns across time (Chauhan & Sehgal, 2017). The criminal investigation is critical for finding crime hotspots. The term is most frequently used to refer to the entire process of obtaining valuable knowledge from data, of which data mining is a subset. Combining data mining and crime analysis makes it possible to assess whether a region is at risk of crime based on the datasets and attributes provided, including the crime type, location of the crime, and date/time of the incident. There is a wealth of data relating to crime in police station records; information about the specific crime or essential information relating to the crime should be extracted (Hassani et al., 2016). It enables the department to carry out various activities, including patrol deployment, special operations, tactical units, investigations, planning and research, crime prevention, and administrative services.

A subset of data analysis called predictive analytics is being utilized by policymakers to help them make better decisions. By utilizing a machine learning algorithm, this methodology for crime investigation becomes well-known and advantageous (Ippolito & Lozano, 2020). Machine learning algorithms such as k-NN, Naïve Bayes, and decision trees

were used in classifying crime in a small amount of data (Wibowo & Oesman, 2019). In an experiment, it is found that the decision tree was a better performing machine learning compared to Naïve Bayes after gaining a much higher accuracy, precision, and recall (Iqbal et al., 2013). It is proven to give a high accuracy; thus, it can still be improved by integrating ensemble methods or application of different feature selections (Aldossari et al., 2020). The decision tree is a reliable predicting algorithm when integrated into computer systems (Ahishakiye et al., 2017). Furthermore, other classification techniques were used to improve the performance of decision trees like regression (Sapin et al., 2021).

Criminal analysis and investigation examine and determine the relationship between crime and offenders (Chen, et. al., 2015). There is a wealth of data relating to crime in police station records; information about the specific crime or essential information relating to the crime should be extracted.

The research study aimed to design an online crime management information system that, using predictive analytics, may contribute to reducing Laguna's rising crime rate. Primarily, the research must accomplish the following objectives: a.) Create a crime prediction model that analyzes index criminal episodes using various attributes from the acquired datasets. b.) Develop a web-based crime information system that handles record management, storage, and retrieval for police stations and administrators covering geographical information and forecasting systems c.) Evaluate the public user view of dynamic Spatiotemporal patterns of criminal activities in the Laguna premises.

LITERATURE REVIEW

Traditional data mining techniques such as association analysis, classification, cluster analysis, and outlier analysis identify patterns in structured data (Han & Kamber, 2006). Classification identifies common characteristics among various criminal entities and groups them into predefined categories. Classification has been widely used to forecast crime patterns, and it can significantly reduce the time required to identify criminal organizations. The technique, however, requires a predetermined classification strategy (Sukanya, Kalaikumaran & Karthik, 2012). Additionally, classification requires complete training and testing data, as a significant degree of missing data reduces prediction accuracy.

In recent times, many researchers have conducted experiments to predict crimes using various machine learning methods and particular inputs. For crime prediction, k-NN, Naïve Bayes, and Decision trees are listed as high-performing algorithms specifically in accuracy. k-Nearest Neighbor (k-NN) is a data classification technique based on the proximity of the data. k-NN makes an explicit prediction on the testing data based on a comparison of k-Nearest Neighbor. To calculate the proximity of data, the Euclidean Distance formula can be used (Thongsatapornwatana, 2016). Bayes is one of the classification methods that can predict the probability of membership in a class. The value of a class in the Naïve Bayes method is independent, which independent of other attributes. It is a simple learning

algorithm that utilizes Bayes' rule together with a strong assumption that the attributes are conditionally independent given the class (Webb, Keogh & Miikkulainen, 2010). A decision tree is a tree that exists in the analysis of problem-solving and alternative solutions mapping which can be taken from the problem. A decision tree can also be called one of the most popular classification algorithms since it is easy to interpret. Decision trees work well with scenarios that produce discrete values. The fundamental advantage of employing a decision tree is that it can be used to analyze and depict complicated decision-making processes in a way that makes them simpler and easier to understand. Typically, decision trees are used to gather the data that informs decisions. A node root (beginning point) is used by the user to start the decision tree. The user solves the leaf nodes under the tree algorithm choice based on this node root. A decision tree with each branch displaying potential decision scenarios and their outcomes is the product of composing the node root and leaf node (Agrawal & Agrawal, 2015).

Abdulrahman and Abedalkhader (2017) compared two types of classifications: the K-NN classifier and the Naïve Bayes classifier. The research used dataset contains incidents derived from the SFPD Crime Incident Reporting system. A total of (878,050) crimes are provided in the training data and the test data for which the categories will be predicted are of the same number data ranges from 1/1/2003 to 5/13/2015. The training set and test set rotate every week. Python was used to explore training data, make regression analysis, and predict categories for test data, to get the best correlation between the features (Date, Pd-District, Address, Day of Week, Description, Resolution, X, and Y) and the target value (Category of Crime). Several trials of different Regression methods were used on the training data by splitting it into two sets; training and validation, both validation and cross-validation were conducted, and the method with the least Log loss was applied to predict the results for the test data. In this research, the two main algorithms were used: k-NN and Naïve Bayes. In the K-NN approach, they applied Uniform and Inverse versions of the technique. For the Naïve Bayes approach, they used three different types; Bernoulli, Gaussian, and Multinomial. They concluded that compared to Naïve Bayes, K-NN has some benefits and drawbacks. As opposed to Naïve Bayes, which can only have linear, elliptic, or parabolic decision borders, k-NN 's decision boundary can assume any shape. Additionally, Naïve Bayes is not effective with correlated features; if correlation rather than marginal distributions serves as the defining feature of classification, NB is not the best option. The lack of an attribute can also mislead Naïve Bayes. One of the drawbacks is that k-NN simply considers distance and does not consider the most crucial characteristics. As a non-parametric model, k-NN cannot reveal any correlations between the distribution of attributes and classes, making it less interpretable than NB. Naïve Bayes simply ignores the attribute of missing data, while k-NN does not appropriately handle missing data. In k-NN, K's value must be adjusted, and an ideal value must be chosen. Another drawback is that k-NN takes longer to process predictions when processing big volumes of data; the speed difference is noticeable.

However, in the study of Sathyadevan and Gangadharan (2014) which created a system that can predict regions that have a high probability of crime occurrence using the Naïve

Bayes algorithm they create a model by training crime data related to vandalism, murder, robbery, burglary, sex abuse, gang rape, arson, armed robbery, highway robbery, snatching, etc. extracting data from different web sites like news sites, blogs, social media, RSS feeds, etc. that gave them 90% accuracy. Another advantage of Naïve Bayes is that it works well for a small amount of training to calculate the classification parameters.

Agrawal and Agrawal (2015) stated in their study that when decision tree and bayesian techniques are compared, though the accuracy of a decision tree is a far better computational time of Bayesian network is low. Hence, when the data set is very large it will be efficient to use NB models. In addition to this, Iqbal et. al. (2013) compared the Naïve Bayesian and Decision Tree for predicting 'Crime Category' for different states in the USA. Using dataset acquired from UCI machine learning repository website which contains a total number of 128 attributes and 1994 instances. The results from the experiment showed that the Decision Tree algorithm outperformed the Naïve Bayesian algorithm and achieved 83.9519% Accuracy in predicting the 'Crime Category' for different states of the USA.

Wibowo and Oesman (2020) tested the various algorithm by implementing three methods, such as k-NN, Naïve Bayes, and Decision Tree, to compare the accuracy. The study was conducted on a data set with a total number of 1,735 crimes data for 3 years, which consisted of 15 attributes, such as Day, Season, Time, TKP, District, Gender of the Victim, Occupation of the Victim, Age, Residence, Sex of the Offender, Occupation of the Offender, Age of the Perpetrators, the Residence, number of the perpetrators, and types of crimes which were classified into 3 classes, such as theft, fraud, and embezzlement. Based on the three algorithms, the accuracy of k-NN with k = 5 was 57.88 percent, with k = 10 was 59.49 percent, with k = 15 is 59.38 percent, with k = 20 was 60.18 percent, and with k = 25 was 61.57 percent. Meanwhile, for the Naïve Bayes algorithm, the accuracy reached 65.59 percent, and the Decision Tree algorithm reached 60.23 percent. In conclusion, the algorithm with the highest accuracy was owned by Naïve Bayes.

The crime pattern and prediction literature examine the relationship between crime and a variety of characteristics, resulting in the development of methodologies for crime forecasting. Most works are devoted to predicting hotspots, and locations of different geographic areas with a high chance of crime. Thematic Mapping is one of the approaches; the map is divided into boundary areas, and offenses are represented as points on the map. After that, the points can be aggregated into geographic unit areas and colored according to the number of crimes committed (Williamson et al., 2001). This technique quickly identifies locations with a high crime rate and allows for an additional examination of the issue by "zooming in" on specific areas.

As a result of various literature reviews and technology, the researchers recognize the value of extracting meaningful information from previously known (historical) crimes and criminals' relationship records to discover the facts surrounding the committed crime and minimize future crime risks and incidents. The information above may benefit the police

authorities in Laguna as they work to reduce criminal activity through resource allocation and proactive crime countermeasures.

MATERIALS AND METHODS

The research undertaking is aligned with the following clusters of disciplines: web development and machine learning, which deal with Crime Incidents in Laguna. To ensure the project's success, various data collection, analysis, and programming tools and techniques are used. The data collection methods used were interview and document analysis. While interviewing important officials, information desk/center officers, and police inspectors used to provide a thorough understanding of the system, observation supports the interview crime reports, and its variables have helped develop the system.

Crime pattern analysis tools have been found to provide valuable insights and knowledge about crime incidence, which can create effective crime prevention operations, particularly in known crime hotspots. Additionally, the current study project aims to assist the police authorities in Laguna in increasing the transmission of crime reports and knowledge throughout the province. The typical knowledge discovery process in databases (KDD) was used to accomplish the goals. KDD is a process for automatically analyzing and modeling huge data repositories through exploratory analysis and modeling. KDD is the systematic discovery of valid, unique, valuable, and visible patterns in vast and complex data sets.

Model Development

Data Collection

Data from police records were obtained with permission from the provincial headquarters of the PNP in Laguna province. The records include details for every crime responded – index and non-index from 2014 to 2017 (n=7,260) such as the date and time of the crime incident, address/location of a crime, type of crime, alert level, and status of the report. This study focuses on index crimes, defined by the Philippine National Police (PNP), involving crimes against persons such as murder, homicide, physical injury, and rape, and crimes against property such as robbery, theft, carnapping/carjacking, and cattle rustling. Target data and Preprocessed data as an output.

Data Preprocessing

Data transformation is a stage in the preparation of data, the process of extracting usable information from data, and entails evaluating and maybe interpreting patterns to determine what constitutes knowledge. Data cleaning was made to the crime data set to make it more analyzable. The dataset used in the experiment contains various items Blotter Entry #, DTG, Place, Law Violated, Victim's Data (Name, Age, Sex, Nat), Suspect's Data

(Name, Age, Sex, Nat), Use Of (Drug, Alcohol, Unknown, Firearms), Remarks, Status, Actions Taken (Barangay, Prosecutor, Court) as reflected on Table 1. The experiment focuses exclusively on factors associated with index crime data from 2014 to 2017. The quality of the mining process's output is directly related to the preprocessed data's quality. It also entailed utilizing the Google Maps API to determine the geographical coordinates associated with each address/location in the dataset. The following properties were generated because of the pre-processing:

Table 1. Dataset Attributes

Attributes	Description	Values
DATE	The exact date the crime occurred	31/12/2016
TIME	The exact time the crime occurred	12:30 AM
ADDRESS	Barangay and City or Town the crime occurred	BRGY SAN ANTONIO SAN PEDRO CITY LAGUNA
VIOLATION	Type of crime occurred	MURDER, CARNAPPING, HOMICIDE, MURDER, PHYSICAL INJURY, RAPE, ROBBERY, THEFT
V_AGE	Victim's Age	48
V_SEX	Victim's Sex	MALE
V_NATL	Victim's Nationality	FILIPINO
S_USE_OF	Suspect's Used Weapon	FIREARM
S_AGE	Suspect's Used Age	35
S_SEX	Suspect's Used Sex	MALE
S_NATL	Suspect's Nationality	FILIPINO
ACTIONS_TAKEN	The action was taken by the officer-in-charge	ARRESTED, AT LARGE, DECEASED, DETAINED, KILLED IN ACTION, RELEASED, TURNED OVER TO DSWD TURN OVER TO WCPD
STATUS	Status of the case	FILED, NOT FILED, PLEASE BE INFORMED...
REMARKS	Special remarks on the case status	UNDER INVESTIGATION, FILED IN COURT, FOR UPDATE, SETTLED

Crime data mining poses concerns about privacy (Kargupta, Liu, & Ryan, 2003) the researchers ensure that specific data like victims' and suspects' data were eliminated as agreed upon the initial agreement upon the receipts of the datasets. The researcher categorized each incident's dates and times by year, month, and day of the week, separating them into distinct columns, and converting time to a 24-hour format. Microsoft Excel was used to visualize the results using graphs and charts to aid in further analysis and interpretation. After a thorough examination, it was discovered that the dataset is missing data for the state and number of crime variables. Two strategies were employed in the

current research endeavor to fill in the missing number of crimes associated with index crimes. To begin, all four fields are checked for empty values. If all four attributes for a specific record are empty, the entire record is considered irrelevant and is removed. From a dataset of index and non-index crimes (n=7,260) to an index crime dataset (n=945).

Building and Training the Model

Choosing a data mining task, classification. Choosing the data mining algorithms Selecting method(s) to be used for searching for patterns in the data, deciding which models and parameters may be appropriate. Searching for patterns of interest in a particular representational form or a set of such representations as classification rules or trees, regression, clustering, and so forth.

Naïve Bayes which was based on Bayes theorem: $P(A|B) = P(B|A) P(A)/P(B)$ applies a probabilistic approach in the dataset (Chauhan & Sehgal, 2017). In this paper, the Bayes theorem can be written as seen in Equation 1:

$$P(y/X) = \frac{P(X|y)P(y)}{P(X)} \qquad \textit{Equation 1}$$

wherein X is set as the element to find in the line of subset which are the attributes of the dataset and represented as x_n The parameters are denoted by variable X. The notation for X is as follows as reflected in Equation 2:

$$\textit{Equation 2}$$
$$X = (x_1, x_2, x_3, \dots x_n)$$

From here *x1, x2, x3,……. xn* represents the attributes such as Date, Time, Address, Violation, etc. After substituting the *X*, it will be expanded using the chain rule and the following will be produced as illustrated in Equation 3:

$$P(y/x_1, x_2, x_3, \dots x_n) = \frac{P(x_1|y) P(x_2|y) \dots P(x_n|y)P(y)}{P(x_1)P(x_2) \dots P(x_1)} \qquad \textit{Equation 3}$$

The dataset would be used to obtain all values. When all the elements in the dataset are substituted, the denominator does not transform and remains constant; hence, it can be omitted, and proportionality is produced as reflected in Equation 4:

$$P(y/x_1, x_2, x_3, \dots x_n) \propto P(y) \prod_{i=1}^{n} P(x_i|y) \qquad \textit{Equation 4}$$

Since the class variable (y) includes more than two outcomes, the classification will be multivariate, and y must be determined with the highest probability as shown in Equation 5:

$$y = argmax_y \, P(y) \prod_{i=1}^{n} P(x_i \,|y) \qquad\qquad Equation\ 5$$

The Naïve Bayes classifier is utilized in crime analysis and is compared to the K-NN classifier (Jantan and Jamil, 2019 & Delima,2019). K-NN is a machine learning technique that is predicated on the occurrence of similar objects nearby. As a result, those characteristics that appear in proximity are deemed true and valuable.
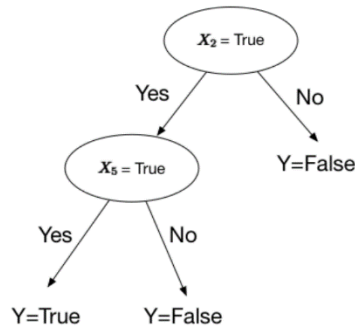

Figure 1. Decision Tree

The data set was divided into two portions for validation and assessment purposes: the training set and the testing set. The training set comprised around 80% of the data set, whereas the testing set comprised the remaining 20%. This procedure was important to assure the classifier model's accuracy and precision and the result is illustrated in Table 2.

Table 2. Train and Test Set

Labels	Train set Number of Instances	Test set Number of Instances
Murder	94	24
Homicide	38	10
Physical Injury	186	46
Rape	87	22
Robbery	94	24
Theft	190	48
Carnapping	65	16
Total	**754**	**190**

This paper also makes use of a decision tree as illustrated in Figure 1, it is well-known for producing results that are comprehensible and legible in natural language. The tree algorithm is represented visually as a flowchart, with external nodes representing the

conclusion and inside nodes representing properties that are deemed to be connected to the class attributes (Ivan et al, 2017). The decision tree diagram is interpreted as conditional statements for pattern acquisition in the dataset. For decision analysis, the combination of all connected internal nodes will be written in an if-else expression.

Prediction Evaluation

The best classifier was picked based on the accuracy, precision, sensitivity, kappa statistics, and specificity of each model as determined by the confusion matrix (Brownlee, 2014). The confusion matrix is a well-known technique for measuring the classification performance of machine learning; it is represented as:

Figure 2. Confusion Matrix

Figure 2 illustrates the Confusion Matric which comprises four (4) distinct attributes: TN, which stands for true negative, FP, which stands for false positive, FN, which stands for false negative, and TP, which stands for true positive. In the confusion matrix, TN and TP denote values that are predicted correctly, but FP and FN denote values that are predicted incorrectly.

Accuracy is expressed as a percentage of properly predicted features; a score of 70% is deemed satisfactory, indicating that the algorithm is capable of accurate prediction. Accuracy can be calculated as shown in Equation 6:

$$\textbf{\textit{Accuracy is }} (\textbf{\textit{TN}} + \textbf{\textit{TP}}) / (\textbf{\textit{TP}} + \textbf{\textit{TN}} + \textbf{\textit{FP}} + \textbf{\textit{FN}}) \qquad \textit{Equation 6}$$

Precision is defined as the number of positive features recognized as positive by the model as reflected in Equation7.

$$\textbf{\textit{Precision is }} (\textbf{\textit{TP}} + \textbf{\textit{FP}}) / (\textbf{\textit{TP}} + \textbf{\textit{FP}}) \qquad \textit{Equation 7}$$

The recall parameter indicates the number of positive features successfully classified by the model as shown in Equation 8.

$$\textbf{\textit{Sensitivity}}(\textbf{\textit{recall}}) = \textbf{\textit{TP}} / (\textbf{\textit{FN}} + \textbf{\textit{TP}}) \qquad \textit{Equation 8}$$

Kappa is a measure of the algorithm's prediction accuracy, ranging from 0 to 1.0, with a score of at least 0.50 deemed satisfactory (Ghoneim, 2019). Kappa Statistics. It computes

the conformity between two rates that classify all n items into m mutually exclusive groups. The value of kappa is defined by Jo and Yoon in 2018.

$$= \frac{P_o - P_e}{1 - P_e}$$
Equation 9

where P_o is the relative observed agreement among rates and P_e is the imaginary probability of chance agreement as illustrated in Equation 9.

The experiments were implemented in the Keras framework using the TensorFlow backend. TensorFlow was utilized as the platform to establish a detection model for traffic classification, and mainly uses the standard data set for related experiments. TensorFlow is an open-source machine learning library developed by Google that can transfer complex data structures to artificial intelligence neural networks for analysis. The classification model chosen will be used to extract and visualize knowledge, as well as to generate customized recommendations based on this knowledge. The established model's viability can be leveraged to display fresh information via data visualization, which can then be used for intelligent reasoning, policy recommendation, and other pertinent objective. The Tensorflow model was generated as json_file which was used in the developed system specifically in the crime map section using React.

System Development

This research was developed with software development methods, namely the agile method. Agile software development is an approach to establishing software with the collaborative efforts of self-regulating and cross-functional teams and the end users for working on both the requirements and solutions (Collier, 2011). The values and principles applied in Agile come from and support the Scrum framework (Larman, 2004). Changes in user opinion about their needs and desires, and predictions that are not under the requirements are the main principles of double recognition in the SCRUM method (Schwaber, 2004)/ There are three core roles in the Scrum framework that work together to produce a product as shown in Figure 3.
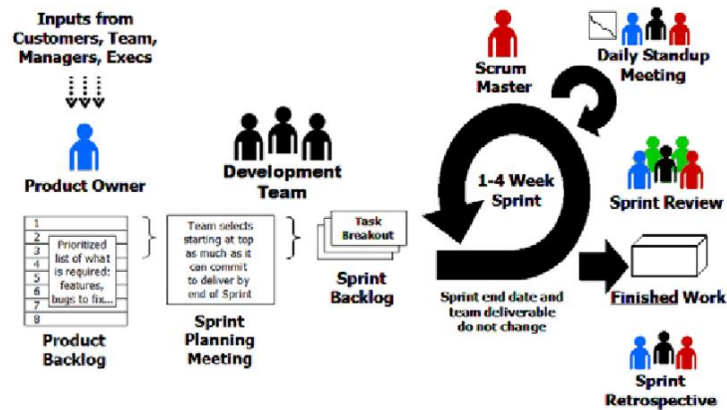
Figure 3. Scrum Model

This work was performed during four project sprints over seventeen weeks. The initial meeting with the PNP-Laguna has been the kick-start of the project. System requirements were specified thru interview, observation, and acquisition of pertinent documents. The main output was a working web-based system, developed, and tested using Scrum and its best practices. In the end, a big data environment was used as a transparent way of fulfilling the needs of crime analysts. The scrum team is composed of a scum master – project leader, 2 programmers, and 3 designers. Regular meetings Upon the development of this project, different ICT tools like geo-mapping, and pattern analysis using the data mining technique.

Software Development Tools

The development of web applications has become a difficult undertaking due to the proliferation of programming and modeling paradigms, notations, and technologies. Integrated development environments (IDEs), relational database management systems (RDBMSs), and other development tools. This research project used various software development tools as indicated in Table 3.

Table 3. Software Development Tools Used

Software	Purpose
ReactJS	Frontend Framework
React Bootstrap	React Components for Faster Development
React Leaflet	Mapping Library for the map
Redux Toolkit	Global State Management
Firebase	Platform used for Authentication and Database
Axios	Handling HTTP requests
Bcryptjs	Cryptography for securing passwords
XLSX	Used to parse excel files
Netlify	Deploying the web application to the internet
Netlify Functions	Serverless functions for Admin tasks

One of them is the ReactJS library, one of the most contemporary web technologies. It has been demonstrated to be the quickest rendering library. It concentrates on the display portion of the MVC paradigm and is extensively used for developing large-scale applications. Because it was designed for internal usage by Facebook, it has shown to be an efficient and quick library in comparison to other technologies.

System Testing and User Evaluation

Alpha testing and user acceptance evaluation survey tools were used to check that the system produced met its specification requirements.

Alpha Testing

At each level of the scrum, alpha testing was undertaken to ensure that all modules operate as intended. This included running functional and non-functional tests to ensure that the system's performance, reliability, and usability were satisfactory.

User Acceptance Evaluation

The survey tool was adapted from the Technology Acceptance Model (Davis, 1989) as the criteria and ISO/IEC 25010 software metrics as the sub-criteria. TAM covers Quality Factors, Perceived Ease of Use, Perceived Usefulness, Attitude Towards Using the Website, Behavioral Intention to use, and experience. Sub-criteria comprise Reliability, Efficiency, Integrity, Learnability, User Interaction Aesthetics, Operability, Usability, Functional Completeness, Suitability, Attractiveness, Operability, Adaptability, Stability, and Suitability. The research focusing human-computer interaction (HCI) evaluation often use discrete multipoint scales 7-point Likert Scale - Level of Acceptability, 7-point Likert Response Scaling (1 = Totally unacceptable, 2 = Unacceptable, 3 = Slightly unacceptable, 4 = Neutral, 5 = Slightly acceptable, 6 = Acceptable, 7 = Perfectly acceptable)

The survey questionnaire underwent several steps of quality assurance to ensure its consistency and trustworthiness in terms of the form and content of the questions. To begin, questions were developed from a variety of survey techniques and tailored to the study's objectives and relevance. Next, the researcher organized a team or group of five faculty members – 4 ICT professionals and 1 development communication specialist – tasked with reviewing and evaluating the questions' appropriateness, given that most things were adapted from foreign circumstances. The researcher integrated all the reviewers' suggestions, criticism, and comments into the final survey questionnaire. Outside of the study's sample, a sample survey of 20 third-year BS Computer Science students was undertaken.

After the validation, 18 statements were produced and 1 open-ended question "*Feel free to write a comment if you have any suggestions regarding the website*." was developed to examine the respondents' recommendations and opinions to improve the website's public access. The researchers demonstrated the website to the 45 randomly selected respondents thru Google Meet and distributed the survey questionnaire thru Google Forms with the website URL to acquire the user behavior and user acceptance of the developed website.

RESULTS AND DISCUSSION

The research aimed to design an online crime reporting management system that, using predictive analytics, may contribute to reducing Laguna's rising crime rate. It focuses on prediction model development, an online crime reporting management system for PNP Laguna, and user acceptance and behavioral intention to use and visit the website providing dynamic spatiotemporal patterns of criminal activities in the Laguna premises. Model Evaluation Result

In this work, three (3) machine learning algorithms were used to develop the model for predicting crime occurrences. Upon evaluating each model, the decision tree and Naïve Bayes both showed good performance in classifying crimes.

Table 4. Classifier Performance

Classifier	Accuracy	Kappa	Precision	Recall	F-score
Naïve Bayes	70.34	0.44	74.52	70.64	72.52
k-NN	64.45	0.16	72.68	633.76	67.92
Decision Tree	84.75	0.64	81.76	83.65	82.69

As the result of the experiment (Table 4) illustrates the classifier performance of Naïve Bayes, k-NN, and Decision Tree. The decision tree showed a very promising result garnering a total of 84.75 accuracy score and an 82.69 score for the f-measure. The model generated from the experiment was integrated into the public access-crime map as seen in Figure 4 which represents the Geographic Information System (GIS) facility of the website which analyzes and displays geographically referenced information specifically crime instances in a unique location.

Developed System Result

The website was developed using Scrum methodology, inputs like functions and features were based on the initial interview and document analysis conducted. Monthly scrums were created and developed comprised of various such as Content Management, Accounts Management, Authentication, Report Management, Generation, and Migration. The website provides different accounts such as system administration (*for system*

management and maintenance), administrator (*for provincial director and authorized access control*), and user (*for local police stations*). The system also provides public access to crime maps and several other websites which do not require an account which this study focuses on user acceptance evaluation.
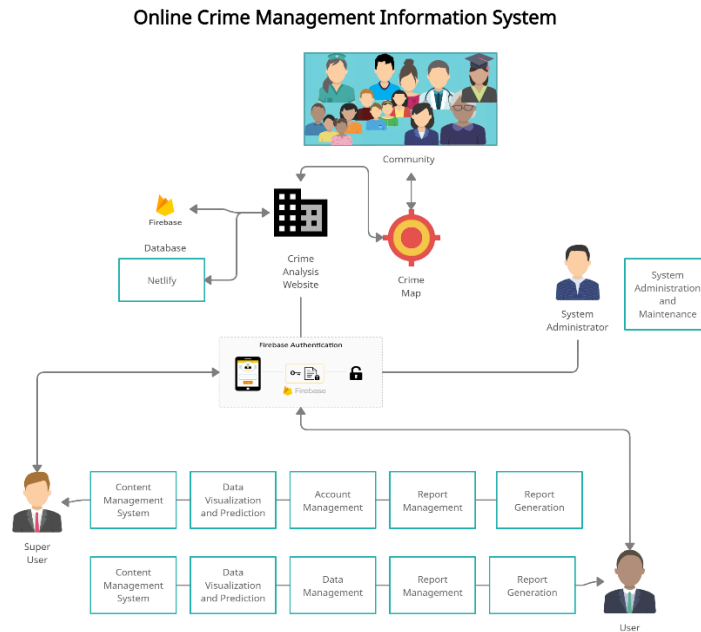

Figure 4. System Architecture of Online Crime Management Information System

The system architecture as seen in Figure 4 was conceptualized and formulated by the project team based on the identified system requirements from the initial coordination. As seen in the figures users are categorized into 3 groups. The System Administrator (for system management and maintenance), the Super User (for provincial director and authorized access control) are illustrated in Figure 5, and the User (for local police stations) is shown in Figure 6 and Figure 7. The system also provides public access to crime maps and crime information and statistics as reflected in Figure 8 which does not require an account in which this study focuses on user acceptance evaluation.

The following figure demonstrates several components of the website dedicated to different users.
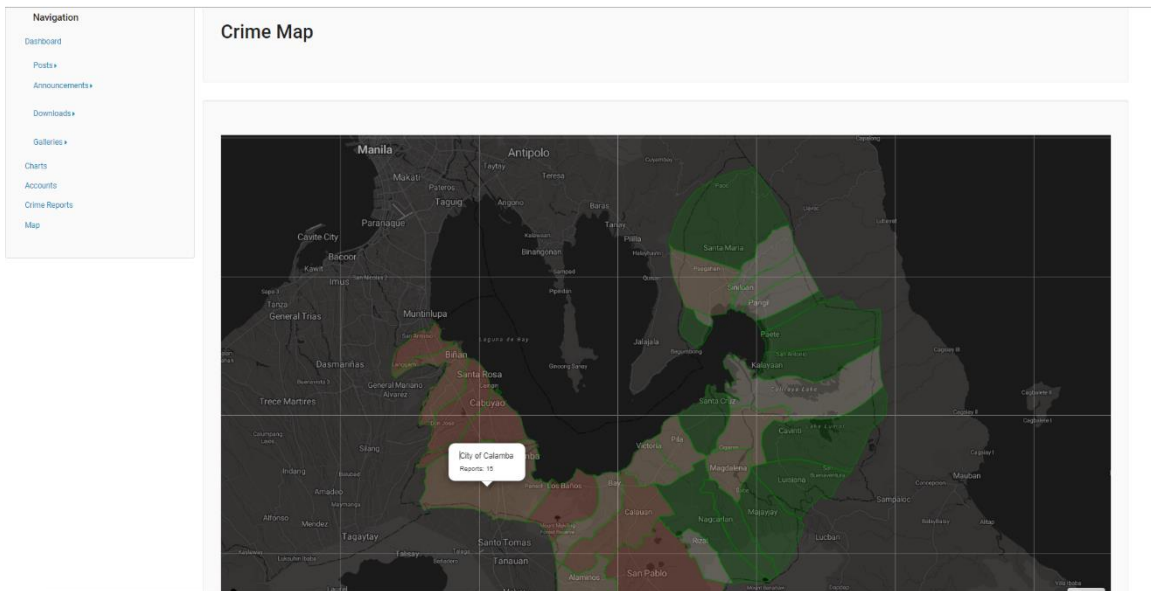

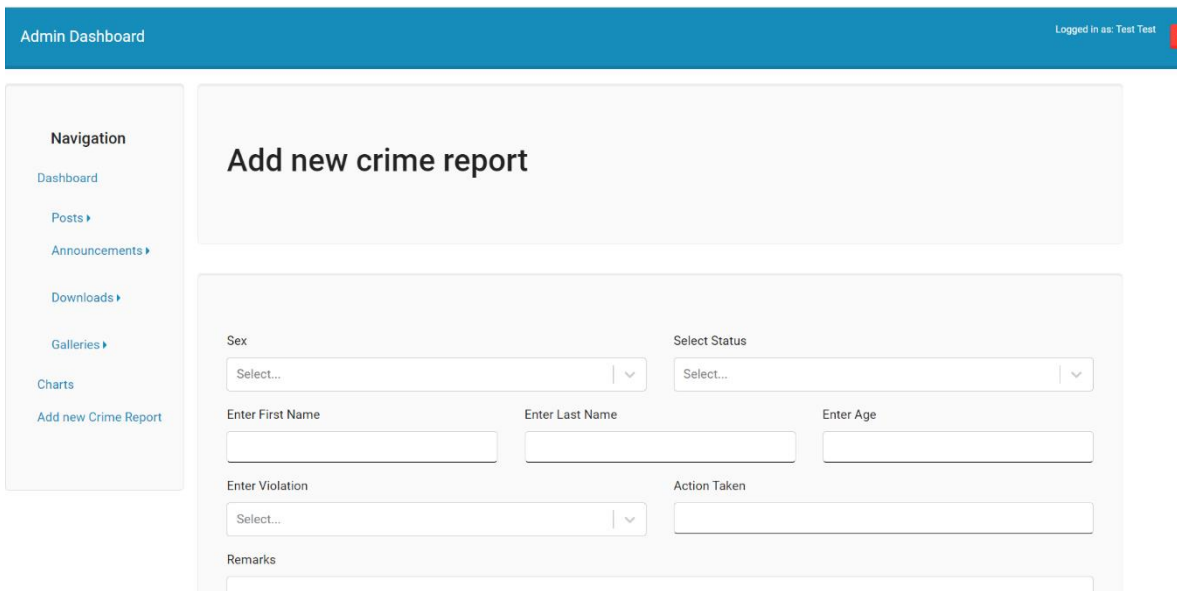*Figure 5.* Screenshot of the Super User Account – Crime Map

Figure 6. Screenshot of the User Account – Data Entry

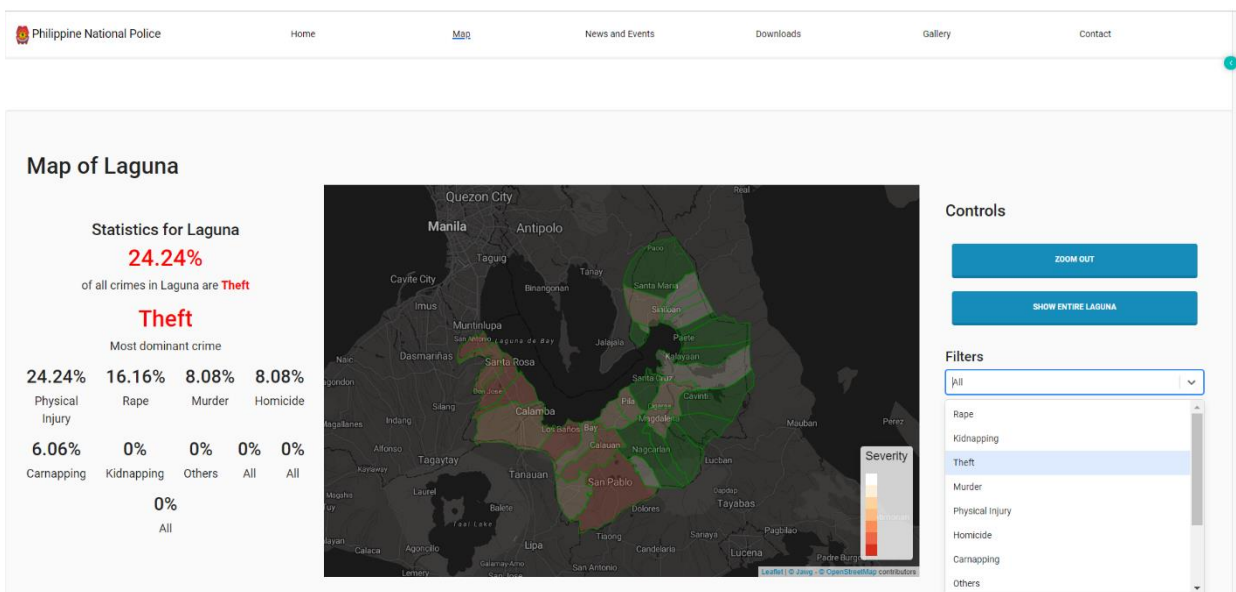*Figure 7.* Screenshot of the User Account – Crime Report Statistics

Figure 8. Screenshot of the Public Access – Crime Map

Figure 8 illustrates the integration of a decision tree-generated model into web development using React. As shown, statistics represented in percentage that the most dominant crimes in Laguna by 24.24% are Theft and Physical Injury. The geo-map shows the severity of Theft in the entirety of the region. From which testing and evaluation of user acceptance are generated.

Testing and User Acceptance Evaluation

The alpha testing was conducted each scrum model was delivered it was evaluated and tested by the researchers and programmer which resulted in a continuous room for improvement. On the public access view of the Crime, map deliverables were tested and deemed to be producing a precise spatiotemporal analysis based on the developed model. However, in the System Administration part specifically, the Super User Account and User account for the Municipality and City Police Station significant findings were found to be improved and enhanced as the project persists. The researchers have several challenges in challenges and issues regarding the development of this component of the website: (1) the COVID-19 pandemic has halted the researchers from physically visiting the PNP Provincial Office which is required in the development of the module (2) The frequent changes in the designated of the Provincial Director since the project started in 2019. However, improvements in the project like support from the Municipal Police Station have already been expressed to acquire the website and assist the researchers to improve the website.

The study focuses on the user acceptability evaluation of the public access view to determine user behavioral intention to use and acceptance of the developed website, as well as to solicit respondents' comments and opinions regarding the website developed by the researcher. The 45 respondents ranged in age from 19 to 35 years and were largely students, instructors of information technology, government employees, and entrepreneurs from Laguna's 4th district – Sta Cruz, Pila, Paete, Los Baos, Pangil, Calamba, Nagcarlan, Mabitac, Siniloan, and Sta. Maria.

Table 4 shows the user behavior and acceptance of the developed website, as reflected in the criteria Attitude Towards Using the Website specifically efficiency which leads to the statement "*Efficiency: The website is reliable in terms of getting the correct spatial and temporal crime information and appropriate crime information for public viewing.*" had the highest mean rated of 6.29 which can be verbally interpreted as Acceptable however the lowest mean rated of 5.91 which falls under the Quality Factors criteria which pertains to the statement "*Reliability: The website has all the necessary functions of the agency that responded appropriately and can be functional in normal operations* which also support on the open-ended response of the respondents "*The website is simple and user-friendly. A little suggestion in the presentation of data. I suggest using a tabular and graphical representation of Data.*"

Table 4. User Acceptance Evaluation Result

Criteria	Mean	SD	Verbal Interpretation
QUALITY FACTORS			
1. Reliability	5.91	0.97	Acceptable
2. Efficiency	6.02	0.99	Acceptable
3. Integrity	6.04	0.95	Acceptable
PERCEIVE EASE OF USE			
1. Learnability	5.98	1.03	Acceptable
2. Understandability	6.11	0.96	Acceptable
3. User Interaction Aesthetics	6.07	1.10	Acceptable
PERCEIVED USEFULNESS			
1. Operability	6.02	0.97	Acceptable
2. Usability	6.16	1.09	Acceptable
3. Functional Completeness	6.04	1.02	Acceptable
ATTITUDE TOWARDS USING THE WEBSITE			
1. Suitability	6.22	0.90	Acceptable
2. Usability	6.20	0.89	Acceptable
3. Efficiency	6.29	0.87	Acceptable
BEHAVIORAL INTENTION TO USE			
1. Suitability	6.07	1.05	Acceptable
2. Attractiveness	6.09	0.95	Acceptable
3. Adaptability	6.11	0.93	Acceptable
EXPERIENCE			
1. Operability	6.02	1.06	Acceptable
2. Stability	5.98	1.01	Acceptable
3. Suitability	6.13	1.04	Acceptable

CONCLUSIONS AND RECOMMENDATIONS

The project establishes a key process for generating an idea that will benefit the community. As a result, the creation of a crime analysis website to provide policy recommendations should be undertaken. It was discovered that predictive crime analytics can significantly assist the PNP in developing strategic plans based on historical data. In terms of crime forecasting, it is found that decision tree algorithms can assist in anticipating trends of location and criminal incidents. Additionally, TensorFlow and Keras were beneficial in developing the predictive model, while React was beneficial in integrating the prediction model into the website. As for the crime analysis website - the crime map for public access user review, which ultimately ended in acceptable, however, the researchers agree with the respondents' comments and suggestions that the crime map should include a variety of graphical representations such as a table ranking the crime

rates from highest to lowest, and a pie graph showing the comparable data of analytics crime per towns and cities to make the system more interesting to any type of user.

Derived from the results, the following recommendations are generated:

1. It is recommended that the PNP-Laguna, municipal police stations, and the LSPU collaborate effectively together to continue improving the crime analysis website, primarily because the PNP-Laguna would benefit the most from the project. The crime analysis website not only covers crime management information systems across police departments for decentralization but also can provide analytics for analytic purposes.

2. It is strongly recommended that Laguna develop and implement an Integrated Crime Information Management System, which will include additional features such as an online crime reporting system for specific localities and an intelligent conversational application that will empower citizens in this time of pandemic the future.

3. Finally, continued experimentation on index and non-index crime datasets to develop intelligent systems capable of forecasting must be explored, as the index crime dataset demonstrated a highly good result in the project.

ACKNOWLEDGEMENT

REFERENCES

Abdulrahman, N., & Abedalkhader, W. (2017). K-NN classifier and Naïve Bayse classifier for crime prediction in San Francisco context. *International Journal of Database Management Systems, 9*(4), 1-9.

Agrawal, S., & Agrawal, J. (2015). Survey on anomaly detection using data mining techniques. *Procedia Computer Science, 60*, 708-713.

Ahishakiye, E., Omulo, E. O., Taremwa, D., & Niyonzima, I. (2017). Crime Prediction Using Decision Tree (J48) Classification Algorithm. *International Journal of Computer and Information Technology*, 6(3), 188-195.

Aldossari, B. S., Alqahtani, F. M., Alshahrani, N. S., Alhammam, M. M., Alzamanan, R. M., & Aslam, N. (2020, January). A comparative study of decision tree and Naive Bayes machine learning model for crime category prediction in Chicago. In *Proceedings of 2020 the 6th international conference on computing and data engineering* (pp. 34-38). ACM. doi:10.1145/3379247.3379279

Amor, N. B., Benferhat, S., & Elouedi, Z. (2004, March). Naive Bayes vs decision trees in intrusion detection systems. In *Proceedings of the 2004 ACM symposium on Applied computing* (pp. 420-424).

Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., & Wu, A. Y. (1998). An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM), 45*(6), 891-923.

Baculinao, R. H., & Ceballos, R. F. (2019). *An analysis of the Location and Type of Index Crimes in the Philippines.* Retrieved from https://psa.gov.ph/sites/default/files/8.6.2

Brownlee, J. (2014). *Classification accuracy is not enough: More performance measures you can use.* Retrieved from https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a general framework and some examples. *Computer, 37(4), 50–56* doi:10.1109/mc.2004.1297301

Chauhan, C., & Sehgal, S. (2017). A review: Crime analysis using data mining techniques and algorithms. *2017 International Conference on Computing, Communication, and Automation (ICCCA).* doi:10.1109/ccaa.2017.8229823

Collier, K. (2012). *Agile Analytics: A value-driven approach to business intelligence and data warehousing.* Addison-Wesley.

Chan, J. (2001). The Technology game: How information technology is transforming police practice. *Journal of Criminal Justice, 1,* 139-159.

Delima, A. J. P. (2019). Applying data mining techniques in predicting index and non-index crimes. *International Journal of Machine Learning and Computing, 9*(4), 533-538.

Ghoneim, S. (2019). *Accuracy, recall, precision, F-score & specificity, which to optimize on?* Medium. Retrieved from https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124

Hassani, H., Huang, X., Silva, E. S., & Ghodsi, M. (2016). A review of data mining applications in crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(3), 139-154.

Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques, 2nd.* The University of Illinois at Urbana Champaign: Morgan Kaufmann.

Harris, C. J. (2007). Police and Soft Technology: How Information Technology Contributes to Police Decision Making (From The New Technology of Crime, Law and Social Control, P 153-183, 2007, James M. Byrne and Donald J. Rebovich, eds.--See NCJ-218026).

Ippolito, A., & Lozano, A. C. G. (2020). Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo. In *ICEIS (1)* (pp. 452-459). doi:10.5220/0009564704520459

Ivan, N., Ahishakiye, E., Omulo, E. O., & Taremwa, D. (2017). Crime Prediction Using Decision Tree (J48) Classification Algorithm. *International Journal of Computer and Information Technology*, 6(3), 188-195.

Iqbal, R., Panahy, P. H., Murad, M. A., Mustapha, A., & Khanahmadliravi, N. (2013). An Experimental Study of Classification Algorithms for Crime Prediction. I*ndian Journal of Science and Technology*, 6(3), 4219-4225.

Jantan, H. & Jamil, A. Association Rule Mining Based Crime Analysis using Apriori Algorithm. *International Journal of Advanced Trends in Computer Science and Engineering, 8*(1.5), 18-24. Retrieved from https://doi.org/10.30534/ijatcse/2019/0581.52019 13.

Jo, H., & Yoon, Y. I. (2018). Intelligent smart home energy efficiency model using artificial TensorFlow engine. *Human-centric Computing and Information Sciences, 8*(1), 1-18.

Kargupta, H., Liu, K., & Ryan, J. (2003, June). Privacy-sensitive distributed data mining from multi-party data. In *International Conference on Intelligence and Security Informatics* (pp. 336-342). Springer, Berlin, Heidelberg.

Larman, C. (2012). *Applying UML and patterns: an introduction to object-oriented analysis and design and iterative development*. Pearson Education India.

Abadin, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M.,..., & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *OSDI '16: Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation* (pp. 265-283). ACM.

Quinlan, J. (2014). *C4. 5: programs for machine learning*. Elsevier.

Reichert, K. (2001). Use of information technology by law enforcement: Promising Approaches to Addressing Crime Series. In the *University of Pennsylvania, Jerry Lee Center of Criminology, Forum on Crime and Justice*.

Roth, R. E., Ross, K. S., Finch, B. G., Luo, W., & MacEachren, A. M. (2013). Spatiotemporal crime analysis in US law enforcement agencies: Current practices and unmet needs. *Government Information Quarterly, 30*(3), 226-240.

Sapin, S. B., Lerios, J. L., Padallan, J. O., Buama, C. A., & Asor, J. R. (2021). Fire incidents visualization and pattern recognition using machine learning algorithms. *Indonesian Journal of Electrical Engineering and Computer Science, 22*(3), 1427-1435. doi:10.11591/ijeecs.v22.i3

Schwaber, K. (2004). *Agile Project Management with Scrum (Developer Best Practices) 1st Edition*. Microsoft.

Schwaber K, & Beedle M. (2001). *Agile Software Development with Scrum*. Prentice-Hall: Upper Saddle River.

Sukanya, M., Kalaikumaran, T., & Karthik, S. (2012). Criminals and crime hotspot detection using data mining algorithms: clustering and classification. *International Journal of Advanced Research in Computer Engineering and Technology, 1*(10), 225-227.

Thongsatapornwatana, U. (2016). A survey of data mining techniques for analyzing crime patterns. In *2016 Second Asian Conference on Defence Technology* (ACDT), 123-128.

Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). *Naïve Bayes. Encyclopedia of Machine Learning, 15*, 713-714.

Welsh, B. C. (2002). Technological innovations for policing: Crime prevention as the bottom line. *Cahiers Polities Studies Jaargang, 2*(1), 129-132.

Zhang, H. (2004). *The optimality of Naive Bayes*. Retrieved from https://www.aaai.org/Papers/FLAIRS/2004/Flairs04-097.pdf

Williamson, D., McLafferty, S., McGuire, P., Ross, T., Mollenkopf, J., Goldsmith, V., & Quinn, S. (2001). Tools in the spatial analysis of crime. In A. Hirschfield and K. Bowers (Eds.) *Mapping and analyzing crime data* (pp. 187-202). Taylor & Francis.

Wibowo, A. H., & Oesman, T. I. (2020, February). The comparative analysis on the accuracy of k-NN, Naive Bayes, and Decision Tree Algorithms in predicting crimes and criminal actions in Sleman Regency. In *Journal of Physics: Conference Series* (Vol. 1450, No. 1, p. 012076). IOP Publishing. doi:10.1088/1742-6596/1450/1/012076.