Received: 06/21/2025; Revised: 06/30/2025; Accepted: 09/06/2025; Published: 10/22/2025

Crime Forecasting Model Using Id3 Algorithm

https://doi.org/10.58870/berj.v10i1.83

Rosemarie M. Perreras
College of Arts and Sciences
San Beda University, Manila, Philippines
rperreras@sanbeda.edu.ph

Jake M. Libed
College of Arts and Sciences
San Beda University, Manila, Philippines
jlibed@sanbeda.edu.ph

Abstract

Crime has been a significant concern in every community, as it brings a negative impact on the economy as well as damages to people's lives and property. A total of 107,538 index crimes were recorded in 2019 (Philippine Statistics Authority, 2021), clearly showing that crime is a major area of concern for the government. Being able to identify significant patterns in crime data would result in a reduction in the crime rate. The study implements the iterative dichotomiser 3(ID3) algorithm to analyze crime patterns in the Philippines. This data analytic approach evaluates significant patterns in identifying key areas that need immediate concern based on data published by the crime research analysis center of the office of the Directorate for investigation and detective management of the Philippine National Police for the period of January to March 2018. Upon the application of the ID3 technique, four significant criteria are used, namely region, robbery incident killed, robbery incident wounded, and shooting incident. The computation for the entropy and gain is obtained from which attribute selection is derived. Among the four attributes from 23 records available, the highest information gain is treated as the most important criterion, and the results show that upon implementing the Shannon entropy calculation, the wounded in a shooting incident obtained the highest information gain and requires immediate attention in region 4A. An evaluation process has been conducted using relative absolute error, mean absolute error (MAE), and F1 score to determine the accuracy of the crime model, and it was able to obtain a value of 95.2827, 0.2964, and 0.905, respectively.

Keywords: Data Analytics; Machine Learning; Crime Analysis; ID3 Algorithm

Introduction

Last November 2000, the United States Conference of Catholic Bishops (USCCB) issued Responsibility, Rehabilitation, and Restoration: A Catholic Perspective on Crime and Criminal Justice, which expressed the stand of the Catholic Church based on the sacred scripture, moral and social teaching, and intended to explain the nature and purposes of criminal punishment. Emphasizing the importance of restitution for victims and promotion of the common good, the Catholic perspective asserts that protecting society and rehabilitating those who violate the law are the primary purposes of punishment. According to the writings of St. Thomas Aquinas, those responsible for imposing order may inflict punishment for the voluntary disturbance of that order, and when a person willingly violates the law, a violation of the civil order and the common good is committed. Thus, civil authorities, being responsible for the civil order, may appropriately administer necessary punishment (Falvey Jr. J. 2004).

A pastoral statement from the Catholic Bishops Conference of the Philippines (CBCP) president, Archbishop Socrates Villegas on the issue of the rising cases of kidnapping in the country in the year 2015, that everyone should be concerned with a crime as it is an offense against human dignity and that battling crime is a duty of every human being in maintaining order in the community.

All United Nations members in year 2015 adopted the 2030 Agenda for sustainable development into which a shared plan for attaining peace and prosperity for people and the planet for today and tomorrow's future from these 17 sustainable development goals (SDGs) are formed as recognizes that the end poverty and other types of deprivations, nations must cooperate with one strategy to improve health, education, reduce inequality, and encourage economic growth while putting an importance on the challenge on climate change and preserving ocean's and forest's (United Nations, n.d.). A global overview of the progress with regard to the implementation of the 2030 Agenda for sustainable development is prescribed in the report of sustainable development goals for the year 2022, which tracks the progress and a thorough analysis of the 17 goals (United Nations, 2022). According to the Report, cascading and interlinked crises are putting the 2030 Agenda for Sustainable Development in grave danger, along with humanity's very own survival. The report highlights the severity and magnitude of the challenges before us. Also, the joining together of crises, dictated by the effect of COVID-19, climate change, and wars, contributes to the impact on food and nutrition, health, education, the environment, and peace and security, and therefore affects all the Sustainable Development Goals (United Nations, n.d.).

Conflict, insecurity, weak institutions, and limited access to justice remain significant concerns to sustainable development, and from here Sustainable Development Goal (SDG) number 16 is formulated to promote peaceful and inclusive societies, provide justice for all, and build effective, accountable, and

inclusive strong institutions at all levels (United Nations, n.d.). In 2019, the United Nations tracked 357 killings and 30 enforced disappearances, one-third (1/3) of the world's population, mostly women, fear walking alone at night, and the global homicide rate falls short of the significant reduction of the 2030 target (United Nations, n.d.).

The Philippines is not exempt from this global crime problem. The country has recorded 68,214 index crimes in the year 2019 and a non-index crime of four hundred eight and nine hundred eighty (408, 980) (Philippine Statistics Authority, 2021). The Philippine National Police (PNP) office of the Directorate for Investigation and Detective Management recorded for the year 2020 the status of the country's security index score, which determines the absence of crime, and obtained a score of 0.7 percent. Also, significant concerns about the identified crimes, such as illegal drug trade, human trafficking, murder, corruption, and domestic violence, have been pointed out (PNP, 2021). The guidelines for the violation of index and non-index crimes is clearly stated in the Republic Act. No. (RA) 3815, which stands as the primary law that governs criminal offenses and their corresponding penalties (Sandy et al., 2019).Maintaining peace and order is a challenging task for the police force as various operational activities, such as area patrol, traffic management, and investigations, are conducted by police operatives to safeguard the welfare of the people and their properties (G. Sarcena & E. Patalinghug, 2021).

Various kinds of research had been conducted to determine the attainment of the United Nation sustainable goals such as the study by (Edralin, D., and Pastrana, R. 2022) that investigates the sustainability initiatives done by companies in the Philippines in attaining good health and Well- Being and Quality Education under sustainable goals numbers 3 and 4 respectively into which they mapped the sustainability programs implemented by the chosen companies that are aligned with their 17 United Nation Sustainable Development Goals. They also determined if their corporate vision-mission statements embed the UN SDGs. Another study by Rada (2022) stresses that language is crucial in the implementation of Sustainable Development Goals (SDGs) and explores a language-based approach to the achievement of SDGs.

Peace, justice, and strong institutions are one of the United Nations sustainable development goals that are devoted to promoting peaceful and inclusive societies as part of sustainable development, providing access to justice for all, and building effective, accountable institutions at all levels (Ackman, M. et al, 2018). As strong institutions, access to justice, and peaceful communities are significant for sustainable development, while in contrast, violence provides a hindrance to the implementation of national development and attaining sustainable development goals (Gill, J. et al. 2021).

Table 1.

SDG 16 Targets and Means of Implementation

Target Description
Significantly reduce all forms of violence and related death rates everywhere
End abuse, exploitation, trafficking and all forms of violence against and torture of children
Promote the rule of law at the national and international levels and ensure equal access to justice for all
By 2030, significantly reduce illicit financial and arms flows, strengthen the recovery and return of stolen assets and combat all forms of organized crime
Substantially reduce corruption and bribery in all their forms
Develop effective, accountable, and transparent institutions at all levels
Ensure responsive, inclusive, participatory, and representative decision-making at all levels
Broaden and strengthen the participation of developing countries in the institutions of global governance
By 2030, provide legal identity for all, including birth registration
Ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements
Strengthen relevant national institutions, including through international cooperation, for building capacity at all levels, in particular in developing countries, to prevent violence and combat terrorism and crime
Promote and enforce non-discriminatory laws and policies for sustainable development

Table 1 above shows the target of Sustainable Development Goal number 16 (SDG 16-Targets and Means of Implementation), which intends to have a safe environment and support the delivery of public services.

Table 2.

Sustainable Development Goal number 16 indicators.

Indicator	Baseline	Latest	Target	Data source Agency
16. 1. Significantly reduce all forms of violence and related death rates everywhere				
16.1.1 Number of victims of intentional homicide by sex and age	12.1 (2016)	4.4 (2021)	Annual reduction of 5% in the proportion to the population 2030	Crime Incident Recording and Analysis System, PNP
16.1.2 Number of murder cases	12,417 (2016)	4,845 (2021)	Annual reduction of 5% in the proportion to the population 2030	Crime Incident Recording and Analysis System, PNP
16.1.3 Monthly average index crime rate	11.3 (2016)	2.8 (2021)	Annual reduction of 5% in the proportion to the population 2030	PNP

In the Philippines, the Philippines Statistics Authority (PSA) is the office in charge of monitoring the Sustainable Development Goals progress. Table 2 (Sustainable Development Goal number 16) below shows the indicator progress for Sustainable Development Goal number 16 in the country.

Table 3.

Advantages and Disadvantages of ID3, C4.5, and CART algorithms.

Algorithm	Advantages	Disadvantages
(Iterative Dichotomiser 3) ID3	• Understandable prediction rules are created from the training data. • Builds the fastest tree • Builds a short tree • Only need to test enough • attributes until all data is classified. • Finding leaf nodes enables test data to be pruned, reducing number of tests. • Whole dataset is searched to create tree	• Data may be over-fitted or over-classified, if a small sample is tested. • Only one attribute at a time is tested for deciding. • Does not handle numeric attributes and missing values.
C4.5	• C4.5 can accommodate both continuous and discrete values. For the algorithm to implement continuous values, it constructs a threshold value and then splits the items into those whose criterion values are beyond the threshold value and those that are lower than or equal to it. • Permits criterion values to be labeled as missing. Lacking attribute values are simply ignored and not employed in gain and entropy computation. • Returns to the ID3 tree once it has been constructed and tries to eliminate branches that do not contribute to prediction by substituting them with leaf nodes	• One of the most important phases in rule generation in C4.5 is its ability to create empty branches. Several leaf nodes with zero values are present and these values contribute either to generating rules or helping to formulate target class for classification purposes instead causing the tree to become more complex. • Generally, overfitting occurs when the method chooses model data with uncommon classification. Normally, C4.5 algorithm implements trees and grows leaf nodes enough to perfectly classify the training dataset and this strategy functions very well in noise data. • Susceptible to noise

Table 3.

Continued.

Algorithm	Advantages	Disadvantages
CART	• CART can easily handle both numerical and categorical variables • CART algorithm will identify the most significant variables and eliminate non-significant ones • CART can easily handle outliers	• CART may have unstable decision tree. Insignificant modification of learning samples such as eliminating several observations and causing changes in decision tree: increase or decrease of tree complexity, changes in splitting variables and values.

Commonly available decision tree algorithms are shown in Table 3. Singh and Gupta identified the advantages and disadvantages of each technique.

Decision tree algorithms such as ID3, C4.5, and the CART algorithm employ the mathematical concepts of formal concept analysis and principal component analysis.

Format concept analysis (FCA) is a field of applied mathematics that is based on a lattice-theoretic formalization of the notions of concepts and conceptual hierarchies. It provides efficient algorithms for analyzing data and discovering hidden dependencies in the data. It also allows the user to visualize the data in an easily understandable way. In Format concept analysis (FCA), data is represented in the form of a formal context, which, in its simplest form, is a way of specifying which attributes are satisfied by which objects. There are some advantages of using Format Concept Analysis (FCA). First is the construction of formal concepts, which extent (i.e., the set of objects) represents an equivalence class of objects sharing the same attributes; and second, the presence of the explicit relationship between equivalence classes. The disadvantage of format concept analysis (FCA) is that the size and complexity of lattices depend on the number of objects and attributes and can be computationally demanding as compared to other clustering techniques.

Principal component analysis (PCA) is a procedure for identifying a smaller number of uncorrelated variables, called "principal components", from a large set of data. The goal of principal components analysis is to explain the maximum amount of variance with the fewest number of principal components. Principal components analysis is commonly used in the social sciences, market research, and other industries that use large data sets. Principal component analysis (PCA) is commonly used as one step in a series of analyses. You can use principal components analysis to reduce the number of variables and avoid multicollinearity, or when you have too many predictors relative to the number of observations. The

advantages of principal component analysis (PCA) are that it involves the lack of redundancy of data, given the orthogonal components. Second, reduced complexity in images' grouping with the use of principal component analysis (PCA). Another advantage is the smaller database representation since only the trainee images are stored in the form of their projections on a reduced basis, and lastly, the reduction of noise since the maximum variation basis is chosen, and so the small variations in the background are ignored automatically. One of the disadvantages of using principal component analysis (PCA) is that the covariance matrix is difficult to evaluate accurately, and lastly, even the simplest invariance could not be captured by PCA unless the training data explicitly provides this information.

In a study conducted by G. Sarcena and E. Patalinghug in 2021, titled "Police Operational Activities and Crime Commission in a City in the Philippines," the focus is on the police operational activities in Pagadian City, Zamboanga del Sur, Philippines. The researchers aimed to assess the extent of police operational activities and examine the association between the crime rate and the scope of police operational activities. Utilizing a quantitative descriptive research method, the study involved 142 active police officers as respondents. The statistical tools employed included frequency count, mean computation, percentage, and regression analysis. The findings highlighted that patrolling and investigation were prominently observed. In terms of index crime, offenses against persons were more prevalent than crimes against property. The researchers recommended strategies such as intensified patrolling in crime hot-spot areas, reinforcement of traffic management practices, acquisition of traffic lights, enhancement of traffic signs, and rigorous implementation of traffic laws and regulations.

Another study conducted by Alvarez, J. et al. 2021 titled "Crime Rate in Zamboanga City: Before and During Quarantine Period," wherein a comparison of the crime rate of Zamboanga City before and during the quarantine period, the covering period is from March 2016 to 2020. Data obtained from the Zamboanga City Police Office was used as secondary data analysis. The findings in their study include the decrease in the volume of crime in the city, as well as Non-index crimes is more prevalent than index crimes in the city. Theft was recorded as the highest index crime volume, followed by robbery and physical injury. Barangays in urban areas in the city recorded the highest crime incidents.

Also, a study by (Maghanoy, 2017) entitled "Crime Mapping Report Mobile Application Using GIS" that created a Geographical Information System (GIS) based Android mobile application to report crime on a given location. The Google heatmap tool is used to report a crime. The respondents of the study were a group of twenty individuals from the vicinity of the University Belt. Statistical treatment of the data involves percentage, weighted mean, and Likert scale.

Lastly, the study entitled "Leveraging Crime Reporting in Metro Manila Using Unsupervised Crowd-sourced Data: A Case for the iReport Framework," into which a mobile application was developed to collect indexed crime reports in Metro Manila through unsupervised crowdsourced data. The main goal of the study is to provide a venue for victims of crimes to report their experiences without having to go directly to police stations, and individuals who have experienced the same offense in the area can link their reports with those previously reported offenses, which refer to the same case.

Statement Of the Problem

Taking into consideration numerous factors in the implementation of a crime analysis model, such as the significant criteria that crime analysts must consider in predicting index crimes, warrants a strong consideration.

Research Objectives

The study aims to utilize the ID3 decision tree technique to analyze crime patterns in the Philippines as an aid towards the attainment of the United Nations Sustainable Development Goal 16 in the country.

Specifically, this study aims to:

1. Implement the ID3 decision tree algorithm to determine significant criteria in
   analyzing crime patterns in the Philippines.
2. Evaluate the ID3 decision tree model using **relative absolute error,** mean absolute error (MAE), and F1 score.

Conceptual Model and Operational Framework

The conceptual model of the study is shown in Figure 1, which implements the use of the Input-Process-Output (IPO) conceptual model, a commonly employed methodology in systems analysis and software engineering, frequently introduced and elaborated upon in numerous programming and analysis texts (Braunschweig, n.d.).

Figure 1.

Conceptual model of the study.

Input

The researchers utilized sample raw data published by the crime research analysis center of the office of the Directorate for investigation and detective management of the Philippine National Police for the period January to March 2018. Table 3 (Crime dataset) shows the recorded index crimes such as murder, homicide, rape, robbery, and theft.

Process

The second phase of the study is the application of the ID3 decision tree algorithm, which identifies significant criteria necessary in examining crime patterns.

Output

The resulting model was able to determine the list of index crimes that need to be prioritized as well as their respective areas based on the input crime dataset.

Table 3.

Crime data set./

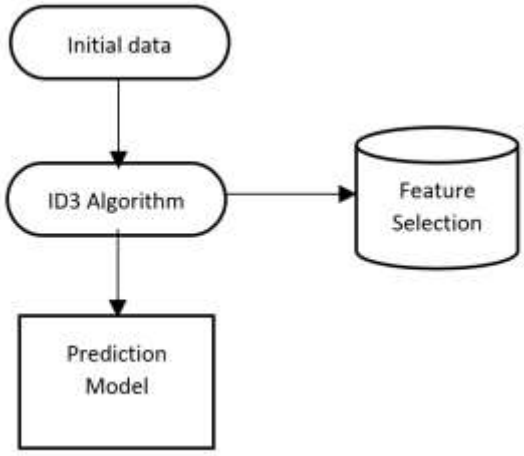
PROs	INDEX CRIME										TOTAL INDEX CRIMES	AVER MONTHLY INDEX CRIME RATE	TOTAL INDEX CRIME CLEARED	INEDEX CRIME CLEARANCE EFFICIENCY	TOTAL INDEX CRIME SOLVED	INDEX CRIME SOLUTION EFFICIENCY
	Against Person					Against Property										
	MUR	HOM	PHY INJ	RAPE	TOTAL	ROB	THEFT	Carnapping		TOTAL						
								MV	MC							
1	79	21	296	84	**480**	69	199	1	30	**304**	**784**	5.03	526	67.09	342	43.62
2	46	19	244	105	**414**	62	126	2	38	**230**	**644**	6.01	550	85.40	364	56.52
3	132	22	393	251	**798**	302	480	11	168	**968**	**1,766**	5.07	1386	78.48	839	47.51
4A	275	45	568	282	**1,170**	372	668	16	200	**1,256**	**2,426**	5.42	1860	76.67	1143	47.11
4B	45	12	107	83	**247**	47	50	0	18	**116**	**363**	3.95	323	88.98	212	58.40
5	97	18	293	104	**512**	191	562	1	83	**841**	**1,353**	7.52	643	47.52	437	32.30
6	114	46	538	94	**792**	156	351	1	32	**544**	**1,336**	5.71	1010	75.60	840	62.87
7	163	39	887	160	**1,249**	500	1770	18	150	**2,442**	**3,691**	16.08	1992	53.97	1724	46.71
8	57	23	276	104	**460**	87	211	2	13	**316**	**776**	5.63	744	95.88	510	65.72
9	95	16	217	88	**416**	168	450	3	67	**690**	**1,106**	9.82	441	39.87	302	27.31
10	82	28	262	98	**470**	114	246	6	74	**446**	**916**	6.29	605	66.05	472	51.53
11	123	18	274	112	**527**	155	276	1	35	**468**	**995**	6.57	567	56.98	428	43.02
12	63	113	237	69	**482**	148	252	0	99	**502**	**984**	6.97	435	44.21	337	34.25
13	52	12	152	47	**263**	70	126	1	21	**219**	**482**	5.98	268	55.60	201	41.70
ARMM	88	7	89	10	**194**	17	16	2	19	**57**	**251**	2.19	56	22.31	32	12.75
COR	17	7	92	39	**155**	39	78	4	12	**133**	**288**	5.39	220	76.39	156	54.17
NCR	226	74	974	299	**1,573**	709	1405	39	213	**2,366**	**3,939**	9.85	2941	74.66	2083	52.88
TOTAL	**1,754**	**520**	**5,899**	**2,029**	**10,202**	**3,206**	**7,266**	**108**	**1,272**	**11,898**	**22,100**	**7.06**	**14,567**	**65.91**	**10,422**	**47.16**
AMCR	**0.56**	**0.17**	**1.88**	**0.65**	**3.26**	**1.02**	**2.32**	**0.03**	**0.41**	**3.80**	**7.06**					

Operational Framework

This section ensures that the research project is well-defined by outlining the specific steps, methods, and criteria used in the study. Figure 2 below shows the conceptual model used in the study. The framework is composed of three (3) major phases; (1) Crime data; (2) Application of ID3 decision tree algorithm; and Prediction model.

Figure 2.

Crime Analysis Architecture.



Crime Dataset

The attributes table

The first phase of the study is to identify attributes. The researchers identified all the attributes that are needed for the crime model process, and it is shown in Table 4 (Attribute names and description).

Table 4.

Attributes name and description.

Attribute	Description
Month	Period covered by crime analysis report
Region	Area covered by crime analysis report
Shooting incident killed	Reported number of shooting incident where casualty occurs
Shooting incident wounded	Reported number of shooting incident where incident of wounded individual occurs
Arrested	Reported number of arrested suspect.

Application of the ID3 Algorithm

The second phase of the study is the application of the ID3 decision tree algorithm. This algorithm utilized the entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous, the entropy is zero (0), and if the sample is equally divided, the entropy has a value of one (1).

Attribute Selection

ID3 used entropy and information gain to construct a decision tree. This determines which attribute is significant to use.

Information Gain.

The reduction in the entropy after training the data, it is divided into an attribute is actually where information gain is centered. Creating a decision tree involves the process of locating an attribute that yields the maximum information gain. It reveals how significant a particular criterion is in a given set, and from here, we will employ this decision to determine a criterion in the nodes of a decision tree. But for us to decide on gain, there is a need to use the idea from information theory, which is the computation of entropy.

Entropy

Entropy is a measure of disorder or uncertainty in a dataset, and it is used to determine the way to split data at each node. A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous, the entropy is zero, and if the sample is equally divided, it has entropy of one.

The Shannon entropy is presented in Equation 1.

Shannon Entropy:

$$\text{Entropy}(S) = -P(p)\log_2(P(p)) - P(n)\log_2(P(n)) \quad \textbf{(1)}$$

P(p) = proportion of positive training examples in S
P(n) = proportion of negative training examples in S

Where:
P(p) is the probability of event ‘p’ occurring.
P9n) is the probability of event ‘n’ occurring.

Step-by-Step Process in computing the entropy and information gain.
Computation for the Entropy of the entire data set.
The initial process we need to perform is to select or construct a target classification class, which is whether the suspect is arrested or not. A detailed computation of this process is shown in Table 5 in order to get the entropy of this data set.

Table 5.

Entropy and data set calculation.

Entropy							
Value	**Total**	**Proportions**	**Y**	**Proportions**	**N**	**Proportions**	**Entropy**
S	23		19	0.826087	4	0.173913	0.666578
Entropy (S) =	**0.666578**						

Computation for the Entropy and Gain for each attribute in the data set.

As reflected in Table 5, a detailed computation is required to obtain the value for the attribute region. There are two values for the attribute region, namely, region 4A and NCRPO, that need to be computed in order to get the entropy and information gain. Based on the computation, the entropy for the region is 0.444380, and the value for gain is 0.222198

Table 5.

Attribute region calculation.

Region							
REGION	**Total**	**Proportion**	**Y**	**Proportion**	**N**	**Proportion**	**Entropy**
4A	12	0.521739	9	0.750000	2	0.750000	0.622556
NCRPO	11	0.478261	10	0.909091	2	0.909091	0.250006
Entropy (Region) =	0.444380						
Gain (Region) =	0.222198						

B.4.3. Select the attribute with the most Information Gain and use it as the root of the tree.

To decide on which value will be on the root node, there is a need to continuously recompute the information gain for all the remaining attributes and obtain the information gain of each attribute.

Table 6 exhibits the outline of the list of calculated criteria. Illustrated on the computation, the criterion shooting incident wounded obtained the highest gain value. Therefore, we will select the shooting incident wounded as our root node for the construction of our decision tree

Table 6.

Computation of highest information gain.

Gain	
Region	0.222198
Robbery incident killed	0.453187
Robbery incident wounded	0.222198
Shooting incident wounded	0.666578

We are now going to evaluate all the attributes after it is computed based on the implementation of the ID3 algorithm. It will also display the predicted output for the historical data. The rules for the application of the ID3 technique are as follows: first, a branch with entropy of zero (0) is a leaf node. Second, a branch with entropy more than zero (0) needs further splitting. Third, is to construct the decision tree and repeat until all nodes are pure (Zero Entropy), and lastly, the ID3 algorithm is run recursively on the non-leaf branches, until all data is classified.

B.4.4. Detailed procedure in constructing a decision tree.

Table 7 illustrates the splitting of shooting incident wounded since it has the highest computed gain value in all the computed attributes, and it also shows the highlighted value on NO because it will later be removed in the constructed decision tree since it obtained an entropy value of zero.

Table 7.

Splitting of the shooting incident wounded.

A split shooting incident wounded							
	Total	**Proportion**	**Y**	**Proportion**	**N**	**Proportion**	**Entropy**
YES	22	0.956522	18	0.818182	4	0.818182	0.473738
NO	1	0.043478	1	1.000000	0	1.000000	0.000000
Entropy (Region) =	**0.453141**						
Gain (Region) =	**0.666578**						

Figure 2 shows that the shooting incident wounded has the highest gain; therefore, it is used as the decision attribute in the root node. A branch with entropy having a value of zero (0) is referred to as a leaf node, and each leaf node signifies a need to decide.

Figure 2.

Root node of the decision tree.

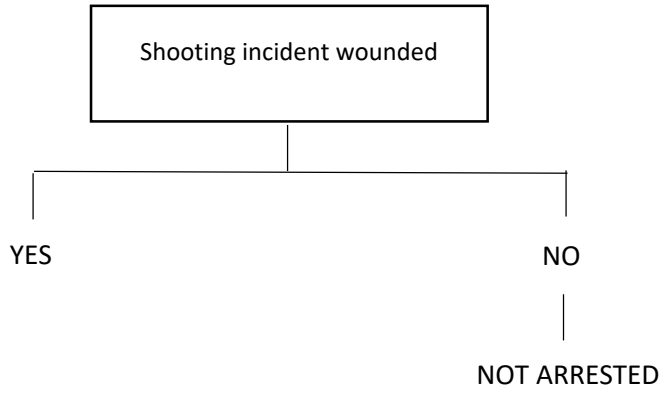


Table 7.

Eliminated dataset on attribute shooting incident wounded.

Month	Region	Shooting incident killed	Shooting incident wounded	Robbery incident killed	Robbery incident wounded	Arrested
January	4A	yes	Yes	Yes	yes	yes
January	NCRPO	yes	Yes	No	yes	yes
February	4A	Yes	Yes	No	No	Yes
February	NCRPO	Yes	Yes	Yes	No	Yes
March	4A	Yes	Yes	No	Yes	Yes
March	NCRPO	Yes	Yes	Yes	Yes	No
April	4A	Yes	Yes	Yes	Yes	Yes
April	NCRPO	Yes	Yes	No	Yes	No
May	4A	Yes	Yes	No	Yes	Yes
May	NCRPO	Yes	Yes	No	Yes	Yes
June	4A	Yes	Yes	No	Yes	Yes
June	NCRPO	Yes	yes	No	No	Yes
July	4A	Yes	Yes	No	Yes	Yes
July	NCRPO	Yes	Yes	No	Yes	No
August	4A	Yes	Yes	No	No	Yes
August	NCRPO	Yes	Yes	No	No	Yes
September	4A	Yes	No	No	Yes	Yes
October	4A	Yes	Yes	No	Yes	Yes
October	NCRPO	Yes	Yes	No	No	Yes
November	4A	Yes	Yes	No	Yes	Yes
November	NCRPO	Yes	Yes	No	Yes	No
December	4A	Yes	Yes	Yes	Yes	Yes
December	NCRPO	yes	yes	yes	yes	yes

Table 7 shows the eliminated dataset on the attribute shooting incident wounded due to the reason of having an entropy value of zero. The need to recompute the information gain for the remaining attributes to update the data set to be used in the crime model.

Table 8.

Updated Entropy and data set calculation.

Entropy							
Value	**Total**	**Proportions**	**Y**	**Proportions**	**N**	**Proportions**	**Entropy**
S	22		19	0.863636	3	0.136364	0.574636
Entropy (S) =	**0.574636**						

Table 8 illustrates the improved data set that is significant to the next process. There is a strong need to recompute our dataset for us to obtain a new value for our entropy. Table 9 also shows the new information gain values using the updated data set.

Table 9.

Re-computation of the highest information gain.

Gain	
Region	0.110056
Robbery incident killed	0.574636
Robbery incident wounded	0.100849

The computation for the Robbery incident killed attribute is shown in table 10 wherein it requires further splitting.

Table 10.

Robbery incident killed attribute.

Split robbery incident killed							
	Total	**Proportion**	**Y**	**Proportion**	**N**	**Proportion**	**Entropy**
YES	6	0.272727	5	0.833333	1	0.833333	0.438391
NO	16	0.727273	13	0.812500	3	0.812500	0.486785
Entropy =	**0.473587**						
Gain =	**0.574636**						

Figure 3.

Attribute robbery incident killed leaf node.

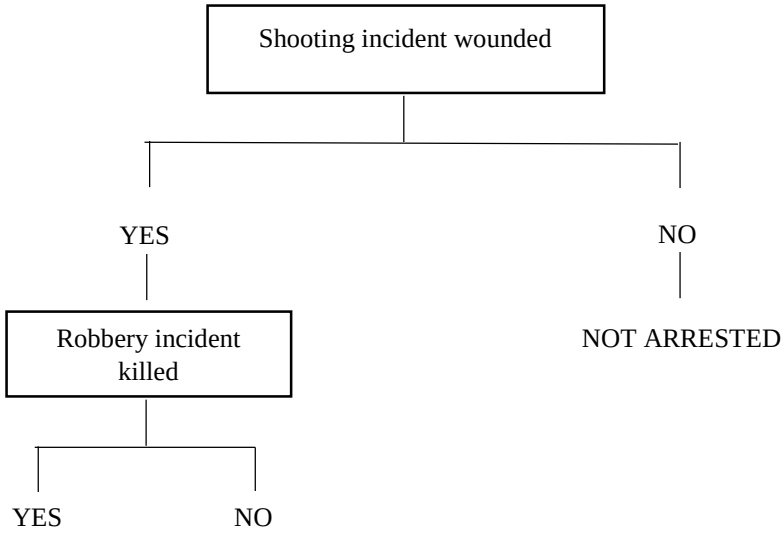


Table 10 illustrates the comprehensive calculation for the criterion of robbery incidents killed. Two values are needed to be computed for us to determine the entropy and information gain. According to the result of the computation of the entropy for the robbery incident, the killed is 0.473587 and the gain value is 0.574636. Also, Figure 3 shows the robbery incident killed attribute having a leaf node.

Research Approach and Procedures

The researchers used action research. Action research employs data collection methods that can be either quantitative, qualitative, or a combination of both. What sets action research apart is its focus on addressing a particular, practical issue and actively seeking solutions to a specific problem. It typically involved the following elements: Identify a problem: Recognize and define a specific issue or challenge that needs attention. Develop a response to the problem: Create a plan or strategy to address the problem based on the research and analysis conducted. Implement the proposed solution: Put the developed plan into action, applying the chosen solution to address the identified problem. Observe the implementation of the solution: Monitor and evaluate the outcomes of the implemented solution, collecting data and feedback.

Results And Discussion

Exhibited in the conceptual model shown in Figure 2 is the crime analysis architecture of the study, which includes the main key features, including crime analysis of data and application of the iterative dichotomiser 3 (ID3) decision tree algorithm.

Table 11.

Summary of ID3 decision tree model.

Mean absolute error	0.2964
Relative absolute error	95.2827 %
F1 score	0.905

Table 11 shows the summary of the ID3 decision tree model of the evaluation procedure. The evaluation utilized Waikato Environment for Knowledge Analysis (Weka) software for the iterative dichotomiser 3 (ID3) decision tree model. Weka is an open-source software that contains a rich collection of machine learning algorithms for data analytics tasks, and it contains important tools for data cleaning, classification, regression, clustering, association rules mining, and visualization. In addition, WEKA performs various data preprocessing steps in data preparation for analysis.

F1 Score is the weighted average of Precision and Recall as presented in Equation 8. F1 score is a data analytic evaluation metric that calculates the accuracy of a model. It combines the precision and recall scores of a model, which also portrays that it accounts for both false positives and false negatives. The higher the value of the precision and recall, the higher the F1-score will be. F1-score ranges between zero (0) and one (1), and the closer it is to the value of one(1), the accurate the model is. The F1-score obtained using the Weka software tool is 0.905.

$$\mathrm{F1} = 2 * \frac{\mathrm{Recall} * \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}} \tag{8}$$

A Detailed summary on the performance measures namely precision, recall and F-measure is illustrated in table 12.

Table 12.

Detailed accuracy.

Precision	Recall	F-Measure
0.826	1.000	0.905

Another significant finding was the score obtained in mean absolute error, having a value of 0.2964. The mean absolute error (MAE) is a metric used to evaluate the performance of the average size of the mistakes in a collection of predictions and is normally measured as the average of the absolute difference between the predicted values and the actual values. It is a measure of the average magnitude of the errors made by the model in its predictions, and is a useful metric for evaluating the performance of a model when the errors are evenly distributed across the data. It is particularly helpful when the errors are symmetrically distributed and there are no significant outliers, since it is not sensitive to the presence of outliers. The mean absolute error states that if the value obtained is close to zero, then the more accurate the model is. Equation 3 shows the formula for the mean absolute error, where n is the number of observations in the data, yi is the true value of the observation, and y is the predicted value of the observation. The vertical bars indicate the absolute value, and the capital Greek letter sigma (Σ) indicates the sum of the differences.

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{3}$$

Where:
MAE = Mean absolute error
Yi = Predictions
Xi = True value
N = total number of items

Relative Absolute Error (RAE) is a way to measure the performance of a data analytics predictive model. It is generally used in machine learning, data analytics, and business management. The Relative Absolute Error is measured in a ratio, comparing a mean error to errors produced by a data analytic model, and it likewise determines the difference between the performance of the predictive model. **The performance of the predictive model is measured as the total absolute difference between the actual and predicted values,** wherein the performance of the predictive model is the total absolute difference between the actual value and the average of all actual values. The **Relative Absolute Error (RAE)** formula is shown in equation 5, and the Weka software tool obtained a score of 95.2827 percent.

$$\text{RAE} = \frac{\sum_{i=1}^{n} |y_i - \hat{Y}_i|}{\sum_{i=1} n |y_i - \bar{y}|} \tag{5}$$

WHERE:
$\boldsymbol{N}$ = The number of total observations
$\boldsymbol{y_i}$ = Actual value
$\hat{\boldsymbol{y}}_i$ = Predicted value
$\bar{\mathbf{y}}$ = The average of the realized values

Conclusion

The researchers were able to provide significant output using the ID3 decision tree algorithm in determining substantial crimes committed. The ID3 technique was iteratively used to determine possible crime scenarios, and its outcome can be used to construct a decision tree that can be used to analyze and identify significant crime patterns. Based on the implementation of the ID3 decision tree technique, region 4A requires immediate attention for the crime shooting incident, which has cases of wounded individuals. Upon implementing the Weka software tool, the crime analysis model achieved a correct classification rate of eighty-two and six 82.6087 percent. Additionally, the F1-score obtained by the model is 0.905, indicating high accuracy in its predictions. Furthermore, the relative absolute error score is 95.2827%, with a mean absolute error score of 0.2964.

Recommendation

In the future, employing a complete and substantial crime dataset that would further enhance the prediction process must be taken into consideration, as well as the implementation of other data analytic techniques in the analysis of crime reports.

References

Afroz, F. (n.d.). *Crime analysis and prediction using data mining – A survey* (pp. 1–6). https://www.academia.edu/29791461/Crime_Analysis_and_Prediction_Using_Data_Mining_CAP_a_Survey

Almuhanna, A. A., Alrehili, M. M., Alsubhi, S. H., & Syed, L. (2021). Prediction of crime in neighbourhoods of New York City using spatial data analysis. *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, 23–30. https://doi.org/10.1109/CAIDA51941.2021.9425120

AL-Huseiny, M. (2015). An e-learning environment based on the Moodle platform for Iraq universities. *Wasit Journal of Engineering Sciences, 3*(2), 5–67.

Alvarez, J. K., Hadjiri, F., Alfaro, R. G., Book, C., Pelisan, C. K., Bechavez, R. C., Molina, R., & Kong, J. (2021). Crime rate in Zamboanga City: Before and during quarantine period. *International Journal of Academic Research in Business and Social Sciences, 11*(3). https://doi.org/10.6007/ijarbss/v11-i3/8935

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Braunschweigh, D. (n.d.). Input-process-output model. Rebus Community. Programming Fundamentals. Creative Commons Attribution-ShareAlike 4.0 International License.

Campanella, P., & Impedovo, S. (2015). Innovative methods for the e-learning recommendation. *2015 5th International Conference on Digital Information Processing and Communications (ICDIPC)*, 312–317.

Calaprice, A. (2010). *The quotable Einstein*. http://www.goodreads.com

Charnock, N. (2016). Teaching methods for balancing chemical equations: An inspection versus an algebraic approach. http://pubs.sciepub.com/education/4/7/2/index.html

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage.

Curtis, H. (1922). The algebraic method of balancing a chemical equation. *Science, 56*(1444), 258–260. http://www.jstor.org/stable/1647136

Edralin, D., & Pastrana, R. (2022). Advancing the pursuit of the United Nations Sustainable Development Goals: Initiatives of selected publicly listed companies in the Philippines.

Fabito, B. S., Lacasandile, A. D., Trillanes, A. O., & Yabut, E. R. (2017). Leveraging crime reporting in Metro Manila using unsupervised crowd-sourced data: A case for the iReport framework. *2017 International Conference on Control, Electronics, Renewable Energy, and Communications (ICCEREC)*, 231–235. https://doi.org/10.1109/ICCEREC.2017.8226681

Filatov, V., Zolotukhin, O., Yerokhin, A., & Kudryavtseva, M. (2019). Personalized adaptation of learning environments. *Proceedings of the International Conference on Advanced Optoelectronics and Lasers (CAOL)*, 584–587.

Gabriel, C. I., & Onwuka, G. I. (2015). Balancing of chemical equations using matrix algebra. *Journal of Natural Sciences Research, 3*, 29–36.

Geddawy, Y. (2019). Adaptive multi-agent assisting framework for personal teaching environment. *Proceedings - Frontiers in Education Conference*.

Gibbs, G. R. (2007). *Qualitative research kit: Analyzing qualitative data*. Sage. https://doi.org/10.4135/9781849208574

Grossi, M., De Elias, M., Chamon, C., & Leal, D. (2018). The educational potentialities of the virtual learning environments Moodle and Canvas: A comparative study. *International Journal of Information and Education Technology, 8*(7), 514–519.

Hamid, I. (2019). Balancing chemical equations by systems of linear equations. *Applied Mathematics, 10*, 521–526. https://doi.org/10.4236/am.2019.107036 https://www.researchgate.net/publication/334396583_Balancing_Chemical_Equations_by_Systems_of_Linear_Equations

Hassan, M., & Hamada, M. (2017). Enhancing learning objects recommendation using multi-criteria recommender systems. *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 62–64.

Hariharan, M. (2019). Teaching style recommender using machine learning. *Proceedings of the 1st International Conference on Advances in Information Technology*.

Intellectual Property Office of the Philippines (IPOPHIL). (n.d.). http://www.ipophil.gov.ph

kotlinlang.org. (n.d.). Kotlin v.16.21 website. https://kotlinlang.org/docs/faq.html#where-can-i-get-an-hd-kotlin-logo

Kularbphettong, K., Somngam, S., Tongsiri, C., & Roonrakwit, P. (2014). A recommender system using collaborative filtering and K-mean based on android application. *Journal of Theoretical and Applied Information Technology, 70*(1), 90–94.

MacCuspie, R. I., & Drake, C. (2014). A framework for identifying performance targets for sustainable materials. *Science Direct, 1-2*, 17–25.

Maghanoy, J. A. W. (2017). Crime mapping report mobile application using GIS. *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, 247–251. https://doi.org/10.1109/SIPROCESS.2017.8124542

Mpungose, C., & Khoza, S. (2020). Postgraduate students' experiences on the use of Moodle and Canvas learning management system. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-020-09475-x

Murad, D., Heryadi, Y., Wijanarko, B., Isa, I., & Budiharto, W. (2019). Recommendation system for smart LMS using machine learning: A literature review. *Proceedings of the 2018 4th International Conference on Computer Engineering and Design (ICCED)*, 113–118.

Panja, B., Meharia, P., & Mannem, K. (2020). Crime analysis mapping, intrusion detection using data mining. *2020 IEEE Technology and Engineering Management Conference (TEMSCON)*, 6–10. https://doi.org/10.1109/TEMSCON47658.2020.9140074

Parker, C., Scott, S., & Geddes, A. (2019). Snowball sampling. *SAGE Research Methods Foundations*. (In Press)

Poulova, P., Simonova, I., & Manenova, M. (2015). Which one, or another? Comparative analysis of selected LMS. *Procedia - Social and Behavioral Sciences, 186*, 1302–1308.

Rada, E. (2022). Language-based approach in achieving Sustainable Development Goals: A qualitative meta-analysis.

Richey, R. (1994). Developmental research: The definition and scope. ERIC. https://eric.ed.gov/?id=ED373753

Saberon, C. B. (2021). Development and pilot validation of E-SCRIBE: An e-learning aid that utilizes algebraic expressions in solving chemical reaction problems (Unpublished master's thesis). The National Teachers College.

Sandy, R., Baculinao, H., & Ceballos, R. F. (2019). An analysis on the location and type of index crimes in the Philippines. https://psa.gov.ph/sites/default/files/8.6.2%20An%20Analysis%20on%20the%20Location%20and%20Type%20of%20Index%20Crimes%20in%20the%20Philippines_0.pdf

Sciences, M. (2021). Crime analysis using data mining techniques and algorithms. May.

Sethupathi, M., & Naresh, R. R. S. (2018). Expert recommendation for crime prediction using data mining techniques. 390–395.

Singh, T., Singla, V., & Bhatia, P. (2016). Score and winning prediction in cricket through data mining. *International Conference on Soft Computing Techniques and Implementations (ICSCTI)*, 60–66.

Srivastava, R., Nigam, S., Srivastava, K., Verma, M., Gorakhpur, M., & Pradesh, U. (2022). Crime prediction and policing. *3*, 243–250.

Sutton, M. (2020, September 21). How to develop a new product (from concept to market). https://www.shopify.com.ph/blog/product-development-process

Syed, T. (2017). A personalized learning recommendation system architecture for learning management system. *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*.

Tiphrak, S., & Kurutach, W. (2015). Ubiquitous computing technologies and context aware recommender systems for ubiquitous learning. *Proceedings of the 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunication and Information Technology*.

United Nations Department of Economic and Social Affairs (UN DESA). (2020). *SDGs learning, training and practice - 2020 edition report*. http://www.sdgs.un.org

Vasquez, J. (2018). The importance of mathematics in the development of science and technology [PDF]. http://verso.mat.uam.es/~juanluis.vazquez/reptmath.pdf

Wonoseto, M., & Rosmansyah, Y. (2017). Knowledge based recommender system and web 2.0 to enhance learning model in junior high school. *2017 International Conference on Information Technology Systems Innovation (ICITSI)*, 168–171.