Malaysian Journal of Computing, 8 (2): 1517-1533, 2023

eISSN: 2600-8238

CYBER FRAUD PROFILING WITH ROUTINE ACTIVITY THEORY USING DATA MINING TECHNIQUES

Sunardi[1], Abdul Fadlil[2], Nur Makkie Perdana Kusuma[3]*

[1]*Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*
[2]*Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*
[3*]*Master Program of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*
[1]sunardi@mti.uad.ac.id, [2]fadlil@uad.ac.id,
[3*]nur2008048034@webmail.uad.ac.id

ABSTRACT

Cyber profiling, as one of the supporting parts of digital forensics, is not only used to record and investigate cybercriminal behaviour. It can also be used to profile victim demographics based on victim characteristics. This study aims to create a cyber-fraud pattern based on a profile created from RAT. This research is expected to be input for internet users in Indonesia, especially IM users such as Instagram, Facebook, WhatsApp, and Telegram. The data collection method in this study uses data mining technology on structured and unstructured data. Structured data was obtained by conducting data mining on the number of cases registered in district courts in Indonesia from January 2021 to January 2022, and the unstructured data was obtained from socio-demographic victims of online crimes. The analysis using the Naive Bayes Algorithm produces a predictive model, which shows the results of online fraud victim profiles based on the weights for each attribute. Cyber-fraud profiling based on RAT with Naïve Bayes Algorithm yields the following findings: Potential Offender Elements: Male, using Facebook, WhatsApp, and Instagram, and crime scene region in Special Capital Region of Jakarta; Elements Suitable Target: Female, using Instagram, WhatsApp, and Facebook, living in the Special Region of Yogyakarta, spending time on the internet more than 8 hours a day, and have more than three IM applications; and Guardianship: Lack of knowledge about Cyber Fraud.

Keywords*: Cyber Fraud, Cyber Profiling, Data Mining, Routine Activity Theory*
Received for review: 28-08-2023; Accepted: 25-09-2023; Published: 01-10-2023
DOI: 10.24191/mjoc.v8i2.23391

1. Introduction

Online fraud is a crime that uses the internet for business and trade purposes, so it no longer depends on actual, conventional business companies. Online fraud is, in principle, the same as conventional fraud. The difference is only in the means of action, using electronic systems (computers, internet, telecommunication devices)(Singh, 2007; Sunardi et al., 2022). Online fraud is included in the crime group of information technology abuse in the form of Computer Related Fraud or can be referred to as Cyber Fraud(Bjelajac et al., 2012; Li, 2020; Sumirat, 2021).

Reep-van den Bergh & Junger (2018) conducted research in Europe by categorizing six types of cybercrime, such as online shopping fraud, online banking/payment fraud, other cyber fraud (such as advanced fee fraud), cyber threats/harassment, malware, and hacking. This research analyzes the percentage based on the number of victims of cybercrimes in Europe based on six categories from 2009 to 2016.

The use of social media (Facebook, Twitter, Google+, and Instagram) and messaging applications (WhatsApp, Telegram, and MiChat) in Indonesia are not strange anymore. All

social media popular in Indonesia require an account registration with an email account and mobile phone numbers (Shrivastava & Jain, 2021). Along with the many attacks from cyber criminals (cyber criminals), several software companies, especially anti-virus and malware makers, are always trying to cover loopholes in existing systems, such as two-step verification facilities (Google mail, Yahoo mail, and steam online).

The use of social media as a medium for promoting goods or services is one of the factors for the increasing use of social media. This utilization applies not only to sellers within an organization but also to individual (private) sellers. The bulk of respondents in a marketing survey (70%) actively use social media for personal sales. Facebook is the most popular social media platform in this regard(Mahmud et al., 2020). The poll revealed that respondents were interested in, trusted, and satisfied with the personal sales-related information offered by the websites. According to marketing studies, customers consider salespeople's social media activity when making a decision (Sianturi et al., 2022). Statistically significant values that reflect the impact of consumer interest, awareness, trust, and pleasure on their purchase decisions on social networks related to personal sales were found.

The increasingly widespread use of social media on the internet also provides opportunities for cybercriminals to take advantage of existing loopholes (Erdoğdu & Koçyiğit, 2021; Shrivastava & Jain, 2021). Starting from cases of fraud via text messages or via WhatsApp, such as pretending to be a family member who asks for phone credit, pretending to be a police officer because the target child is in custody, or pretending to be a doctor that helps the target father, and many more. In addition to using fraud through messages, there are also several crimes in cyberspace by sending links via direct messages on social media (Instagram or Facebook) that can make criminals take control of the victim's account.

Humans are the weakest elements of a security system (Stajano & Wilson, 2011). Alzubaidi (2021) conducted research that measured the level of cyber security awareness in Saudi Arabia, in terms of cyber security, awareness levels, and incident reporting, through an online questionnaire with 1.230 participants. The research shows that 31.7% use public Wi-Fi to access the Internet, 51% use their personal information to create passwords, 32.5% have no idea about phishing attacks, and 21.7% have been victims of cybercrimes. In comparison, only 29.2% reported a crime, which reflects the level.

In Indonesia, online account security is still not important, whether email accounts or other social media accounts, as seen from passwords that are not unique, not intelligent, and unsafe. Although some email providers provide two-step verification for signing in, in Indonesia, it is still unclear whether this feature has become something that is considered important by social media account owners. The lack of cyber security knowledge is one of the gaps that can be used as a gap for cybercriminals, especially with social engineering (Mamade & Dabala, 2021). Online criminals often take advantage of this loophole to capture their victims. An Evaluation Model for Online Crime already exists. However, the details for each factor have not been optimized. They have never been used for case studies from the victim side in Indonesia. Social engineering itself has not received special attention for internet users.

To begin with, Figure 1. is an example of fraud by exploiting human weaknesses via SMS in Indonesia. The perpetrator will pretend to be the victim's parent and try to trick the victim into topping up credit to the number provided by the perpetrator (impersonation). Figure 2 also shows online fraud in the same method using messaging application in the Facebook social media feature. Figure 3 is an online fraud on Instagram pretending to be the admin of Instagram that sends phishing links. This fraud often occurs, which results in the theft of Instagram accounts. The perpetrator has full access to the Instagram account caught in a phishing trap.

Figure 1. Online Fraud method by pretending to be the victim's mother asking for credit via SMS.

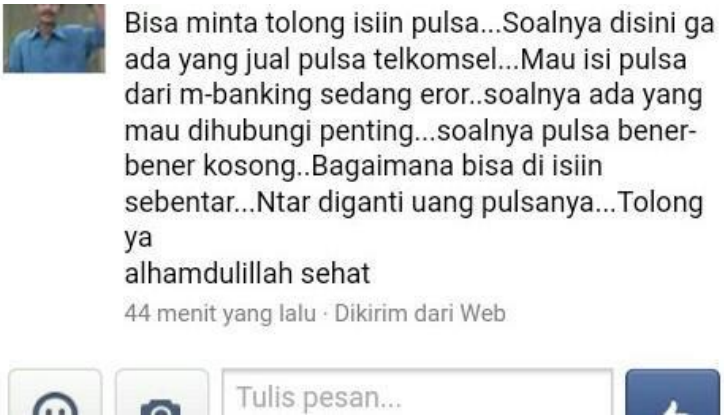

Figure 2. Online Fraud method by pretending to be the victim's relative asking for credit via Facebook Messenger.

Apart from taking advantage of the psychology of the account owner by impersonating using IM in terms of fraud or online buying and selling transactions, many criminals use other people's photos or fake testimonials to convince victims. Examples of cases in the Case Tracing Information System in Indonesia with Case Numbers: 245/Pid.Sus/2021/PN Smn and 350/Pid.Sus/2021/PN Smn reveals that Cyber fraud in buying and selling transactions is also done using IM. Perpetrators post photos/pictures of merchandise that do not exist and wait for potential victims interested in the goods being sold to contact the perpetrator. In cases that go into court, the perpetrators post on Facebook or Instagram homepages and then communicate using WhatsApp.

Marshal (2009) developed an approach to cybercrime, namely a model of online crime evaluation to create cyber profiling. Testing of the model gives results with parameters that indicate the probability of a cyber-attack's success through several elements. These 6 (six) factors can estimate the likelihood of an attack from a cyber-criminal in the evaluation model for an online crime: Criminal Expertise, Victim Expertise, Criminal Guardianship, Victim Guardianship, Nature of the Attack, and Criminal Freedom.

Cyber profiling, as one of the supporting parts of digital forensics, aims to provide information about the perpetrators' characteristics, motives, characteristics of victims, and backgrounds. Cyber profiling is not only used to record and analyze cybercriminal behavior. However, it can also be used to profile victim demographics based on victim characteristics such as gender, age, knowledge of information technology and computers, education level, area of origin, and occupation.

Saroha (2014) states that profiling cyber criminals will help the authorities narrow the scope of their search so that they become more focused on intensively searching other available sources. Technology is indeed the primary defense against cyber-attacks. A better understanding of psychological, criminological, and sociological aspects can provide input for protection efforts and catch cyber criminals before the distance gets too far.

Criminal profiling does not necessarily provide the specific identity of the perpetrators of crimes (K. Sindhu & B. Meshram, 2012; Sebastian et al., 2023; Tompsett et al., 2005). The application of this method is only show the types of people who are most likely to commit crimes by focusing on specific behavioral and personality characteristics [11]. Profiling is generally done for criminals, but it can also be used for profiling crime victims. Suppose profiling perpetrators of crime aims to make it easier to catch perpetrators. In that case, profiling of victims of crime is intended to make it easier to target information dissemination and carry out prevention efforts (Ahmad & Thurasamy, 2022; Hawdon et al., 2017; Leukfeldt, 2014).

A crime victim profile can be defined as a series of approaches and techniques used to predict the characteristics of an unknown perpetrator through investigating and analyzing evidence obtained from a crime scene. By analyzing crime scenes, investigators aim to understand the perpetrator's personality, demographic and behavioral characteristics. The characteristics obtained from the crime scene can be used to identify the behavior patterns of the perpetrators and create a psychological portrait of the perpetrators. Victim profiling is generally used to determine cause-and-effect relationships between crime scenes, victims, and witnesses perpetrators'. This technique is widely used in crime scenes where the perpetrator's identity is unknown and the types of crimes are severe such as murder and rape (Stajano & Wilson, 2011).

One of the popular techniques to create a criminal profile is the Routine Activity Theory (RAT). The theory of routine activity, first formulated by Lawrence E. Cohen and Marcus Felson and later developed by Felson, is one of the most cited and influential theoretical constructs in criminology and crime science (Schreck, 2017). In contrast to criminal theory, which focuses on the criminal and the psychological, biological, or social factors that underlie criminal acts, routine activity focuses on the study of crime as an event, highlighting its relationship to space and time and emphasizing its ecological nature and implications.

Yar (2005) explores the RAT developed by Marcus Felson and others, the extent to which theoretical concepts and etiological schemes can be transferred to crimes committed in 'virtual' environments. Substantively, conventional crime theory's core concepts can be applied to cybercrimes. There are nonetheless significant differences between 'virtual' and 'terrestrial' worlds that limit the theory's usefulness. These differences provide qualified support for the notion that 'cybercrime' does represent the emergence of a new and distinctive form of crime. Choi (2008) empirically assessed the victim model of computer crime by applying the Routine Activities Theory. The survey was conducted on 204 students to collect data to test the model. The findings of this study provide empirical support for the Routine Activities Theory component by describing patterns of computer crime victimization.

The basis for making online fraud patterns uses Routine Activity Theory (RAT) as the basis for describing criminal events through three important elements centered in space and time in daily activities, namely (Agustina, 2015; Bock et al., 2017; Choi, 2008; Kigerl, 2012; Ngo & Paternoster, 2011):

1. Potential offenders with the capacity to commit crimes,
2. Suitable targets or victims;
3. The absence of a capable guardian.

Crime pattern theory highlights the spatial ties that link crime, targets, and movement patterns of offenders whose routine activities occur at places and times where there is a greater likelihood of committing illicit acts (Goni, 2022). Perpetrators commit crimes near areas where perpetrators spend most of their time (home, work, schools, shopping, and entertainment venues) and around routes connecting perpetrators to victims. Awareness of the space around the actor is determined by the activities carried out in the past and the conditions under which the activity will be placed in the future. It is necessary to understand the pattern of daily life and the movement of criminals to comprehend the distribution pattern of crime.

Over the years, routine activity theory has significantly impacted criminology and received important empirical support. This approach closely relates to crime analysis, prevention, rational choice, and crime patterns. It has been applied to problem-oriented situational prevention and analysis strategies with significant effectiveness.

Cybercrime will be patterned based on RAT, combined with classification techniques in data mining to perform profiling based on cybercrime attack patterns, victim socio-demography, and media used as cybercrime tools. Data mining can be used to find interesting patterns and correlations in data, which can then generate knowledge. Data mining has recently become relevant in tourism because of its potential to uncover undiscovered patterns in large data sets and, unlike other statistical approaches, its ability to examine non-linear correlations in the analyzed data. Data mining has fewer assumptions about data quality than other statistical approaches because data may be incomplete, noisy, redundant, and dynamic (Ritonga & Muhandhis, 2021).

The patterns in this data mining can be statistical; an example, the unemployment rate can be derived and predicted using data mining. Correlation can also be used in the realm of machine learning. For example, using data mining into machine learning programs to predict customer behavior (Hassan & Mirza, 2018). Data mining can also create knowledge-based systems to predict cybercrime patterns (Michael, 2020).

Classification is one of the methods in data mining that aims to target categories or classes. Classification is a method in data mining commonly used to predict individual results based on the input given. The purpose of classification is to accurately predict the target class for each case in the data. The algorithm processes a training set containing attributes and their respective outcomes to predict outcomes. The classification algorithm analyzes the input and generates predictions, such are Naïve Bayes, Decision Trees, and Random Forests. Palaniappan et al. (2017) research focuses on helping banks improve the accuracy of their customer profiles through classification and identifying groups of customers with a high probability of subscribing to long-term deposits. In this study, three classification algorithms were used, namely Naïve Bayes, Random Forest, and Decision Tree, to measure the percentage of accuracy, precision, and recall rate. Classification helps predict customer profiles and increase telemarketing sales. According to Sunardi et al. (2023), Naïve Bayes has the best accuracy in creating cybercrime victims' profiles, with 77.3% than Decision Tree and Random Forest.

Aimran et al. (2022) compared six predictive models: Decision Tree (C5.0 and CHAID), Logistic Regression (Forward et al.), and Artificial Neural Network (Multi-Layer et al. Function). Among these models, the Decision Tree (C5.0) demonstrated the highest performance in accurately classifying divorce among Malaysian women, achieving an accuracy of 77.96%. The Artificial Neural Network (Multi-Layer Perceptron) and Logistic Regression (Forward Stepwise) models followed with accuracies of 74.68% and 67.89%, respectively. Shahira Pisal et al. (2022) presents machine learning algorithms for life expectancy based on the Asian population dataset. Comparisons are made between tree classifier models, namely, J48, Random Tree, and Random Forest.

This study aims to create a cyber-fraud pattern based on a profile created from RAT. This research is expected to be input for internet users in Indonesia, especially IM users such

as Instagram, Facebook, WhatsApp, and Telegram. With this research, it is hoped that it can map the need for socializing online crime to IM users to reduce the number of online crimes in Indonesia.

2. Methodology

To begin with, the data collection method in this study uses data mining technology on structured and unstructured data as shown on Figure 3. Structured data was obtained by conducting data mining on the number of cases registered in district courts in Indonesia from January 2021 to January 2022. At the same time, the unstructured data was obtained from an online survey using Google Forms to obtain socio-demographic victims of online crimes.

Structured data is obtained by taking reports from the Case Tracing Information System from all regions in Indonesia from January 2021 to March 2022. Data collection uses a probability sampling technique based on the criteria for Information and Electronic Transactions cases, particularly Cyber Fraud. The data collected is adjusted to the needs of the RAT, such as Gender, Region, and Type of Cyber Fraud. The researcher can find only this data with certainty in Case Tracing Information System in Indonesia. At the data collection stage from Case Search Information Systems, the researcher used the help of an add-on from Google Chrome, namely Data Scrapper, to extract the things available in Case Tracing Information System in Indonesia.

Beranda	Perdata Umum	Perdata Khusus	Pidana	Pidana Khusus	Jadwal Sidang	Laporan	Delegasi			
						Yudianto				
10	78/Pid.Sus/2022/PN Yyk	13 Apr 2022	Informasi dan Transaksi Elektronik	Penuntut Umum: 1.DANANJAYA WIDIHARSONO SH KN MH 2.ARIF MUDA DARMANTA SH MH Terdakwa: 1.OLEKSANDR CHUIKO 2.MARYNA KAHALNYTSKA	Minutasi	97 Hari	[detil]
11	345/Pid.Sus/2021/PN Yyk	22 Dec 2021	Informasi dan Transaksi Elektronik	Penuntut Umum: SUYATNO, SH. Terdakwa: Faisal Umar Firmansyah Bin Farid Firmansyah	Minutasi	57 Hari	[detil]
12	243/Pid.Sus/2021/PN Yyk	26 Aug 2021	Informasi dan Transaksi Elektronik	Penuntut Umum: SUYATNO, SH. Terdakwa: 1.Dimas Hardiansyah Als Pak Dhe Bin Tori Sudibyo 2.Kiki Alfiani Binti Wawan Subandi	Minutasi	53 Hari	[detil]
13	242/Pid.Sus/2021/PN Yyk	26 Aug 2021	Informasi dan Transaksi Elektronik	Penuntut Umum: FADHOLY YULIANTO, SH.MH Terdakwa: Yuci Harucl Als Uci Farantika Binti Agus Wahyudo	Minutasi	53 Hari	[detil]
14	323/Pid.Sus/2020/PN Yyk	16 Dec 2020	Informasi dan Transaksi Elektronik	Penuntut Umum: NURHAYATI, SH Terdakwa: NATHANAEL BUDHI SUSILO Alias CUN CUN Bin AGUS PURWANTO	Pencabutan Perkara Banding	75 Hari	[detil]
15	287/Pid.Sus/2020/PN Yyk	18 Nov 2020	Informasi dan Transaksi Elektronik	Penuntut Umum: ARI MARTINI, SH. Terdakwa: EKO GIAT DARYADI Bin DJUWANDI	Minutasi	63 Hari	[detil]

Figure 3. Case Tracing Information System in Yogyakarta, Indonesia.

Data collection for unstructured data is done by distributing surveys. The researcher did this survey based on research data from Alzubaidi (2021a), which was later adopted for the Indonesian region. This research distributes a questionnaire containing questions about the socio-demographics of cyber fraud victims.

The survey includes questions about age, gender, region, education level, occupation, instant messenger used, gadgets used, and Internet use duration in one day, cyber-security, and cyber-crime knowledge. The questionnaire was created using Google Forms and broadcasted online to active internet users in Indonesia with the snowball sampling technique from January 2021 to March 2022.

The researcher limited participants to Indonesian nationals older than 18 and only one submission per participant was permitted. The link is shared through WhatsApp, and to answer the surveys, login was required using a Google account to prevent duplication.

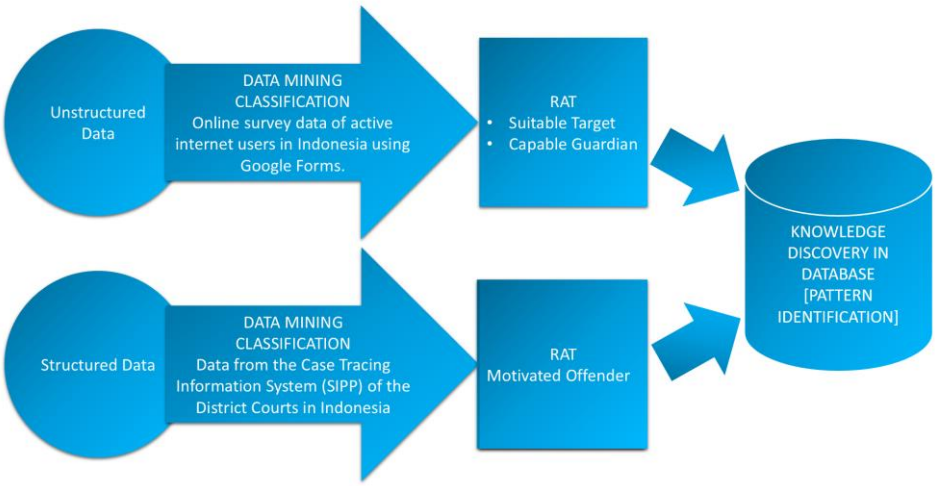

Figure 4. Research flow.

Data collection on Case Tracing Information System is comprehensive in all provinces in Indonesia. The data is then tabulated into a worksheet with the attributes Type of Crime, Media, Gender, Mode, City/District, and Province. One thousand forty data from Case Search Information Systems can be used in all electronic transaction cases in this study. Researchers categorize cases of Electronic Information and Transactions, namely: Fake News, Content Violating Decency, Pornographic Content, Treason, Extortion, Defamation, Online Fraud, Online Narcotics Sales, Unpleasant Acts, Online Gambling, Piracy, Online Prostitution, Skimming, and Hate Speech.

In this study, the data is processed first by cleaning to remove data that cannot be used for the analysis stage. Inappropriate or incomplete records are removed. In the data from the online survey, only three were found to be inappropriate, so they were omitted.

This study uses two data tabulated into worksheets, namely data from Case Tracing Information System and data from surveys as shown in Figure 4. The two data are then entered into RapidMiner to be continued into the model design process according to the method used, namely the Naïve Bayes Algorithm, Decision Tree and Random.

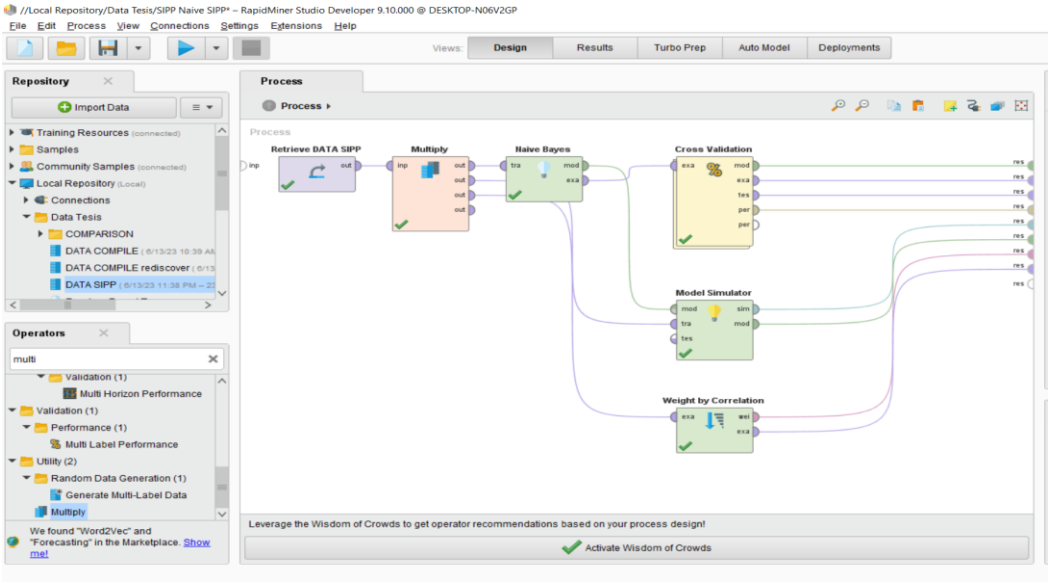

Figure 5. Naïve Bayes model using RapidMiner.

Figure 5 shows the design of the three algorithms is similar for data from Case Search Information Systems and data from GoogleForms, which differs only in the part of the data used. The design uses the Cross Validation operator, which is used to automate training data and testing data. This operator is also helpful for calculating the performance of each algorithm based on accuracy, precision, and recall. By adding the simulator model operator and Weight by Correlations, this design can simulate the algorithm prediction for the class being tested based on the existing attributes based on the weighting of the existing attributes.

Model testing is done by running the system for each data with three different algorithms. Testing is carried out based on the performance of each algorithm, namely the performance of accuracy, precision, and recall. This performance is generated using the Confusion Matrix theorem. At this stage, the researcher simulates the model generated from the best algorithm where the results can be seen in real-time and checks whether the data changes in the model behave according to the test results. Online profiling of perpetrators and victims of fraud is carried out by adopting parameters based on the results of the RapidMiner simulator model.

In addition, both datasets were first processed using RapidMiner software. RapidMiner is a pre-processing tool at this stage to clean and delete data that is considered unsuitable for analysis. Suppose the data obtained from the two datasets do not meet the requirements of the classification model. In that case, RapidMiner advises whether the researcher will still include the data in the classification process.

Furthermore, RapidMiner performs classification and text-mining processes on structured and unstructured data. This process is carried out separately. Classification and text mining on unstructured data is used to complete the data in two elements of the RAT: Suitable Target and Capable Guardian. Meanwhile, classification and text mining on structured data is used to complete the data in the RAT, Motivated Offender element.

In the last stage of this research, the processed data is then analysed based on the attributes needed in profiling. The information that has been obtained is calculated as the weight of each index. The final stage is to evaluate the data mining results and whether the applied model profile can meet the goals. The profiling generated from the data mining process can be used as a recommendation to determine strategies to reduce the number of online frauds in Indonesia.

3. Results and Discussion

Analysis of unstructured data produces variables that can be used as elements of the RAT, namely suitable target elements. Targets are categorized based on socio-demographics by adopting based on research conducted by Alzubaidi. These data become the values to make victim's profiling with the RAT.

The first analysis is based on a survey of 1,587 participants actively using IM. In the pre-processing stage of data mining, the data is first cleaned to remove inappropriate or missing data. 1220 respondents were found to have been victims of cyber-fraud via IM, and 367 had never been victims. These characteristics are divided into age, gender, education, occupation, region, time spent using internet use in one day, number of installed IMs, gadgets, and IM media used.

Accessible targets can include a person, object, or place. In this study, the target is victims of online fraud who suffer losses. Analysis of unstructured data obtained results based on socio-demographic characteristics of online crime victims, such as age, gender, education, regions, time spent accessing the internet in a day, and gadgets and applications used to use the internet.

Table 1 shows the results of a survey of 1220 participants who have been victims of online fraud in the analysis using text mining with RapidMiner RapidMiner to extract social media information. The top three social media that victims often use are Instagram, WhatsApp, and Facebook. Table 2 shows that victims often use devices such as smartphones (Android and iPhone).

Table 1. Instant Messenger That Often Used By Cyber Fraud Victim.

Instant Messenger	Number of Instant Messenger used by Victim
Instagram	699
WhatsApp	691
Facebook	483
Telegram	339
TikTok	217
Twitter	141
Line	96
Others	21
MiChat	17
SnapChat	4

Table 2. Gadget That Often Used By Cyber Fraud Victim.

Gadget	Number of Gadget used by Victim
Smartphone (android/iphone)	726
Laptop	157
Computer	36
Tablet	21
Others	10

The performance of the Naïve Bayes model is based on the percentage of Accuracy, Classification Error, Precision, and Recall. Table 3 shows that the Naïve Bayes and Decision Tree have Accuracy and Precision values of 77.3%, and Recall is 100%. The results also show that Naïve Bayes has a 23.7% on Classification Error. With an accuracy percentage of 77.3%, the prediction results from this Naive Bayes algorithm and Decision Tree can be used as data for the RAT element, the Suitable Target.

Table 3. Performance on Predicting Cyber Fraud Victim Using Confusion Matrix.

Model	**Performance**			
	Accuracy	**Classification Error**	**Precision**	**Recall**
Naïve Bayes	77.3%	23.7%	77.3%	100%
Decision Tree	77.3	23.7	77.3	100
Random Forest	76.8	23.2	77.2	99.4

In structured data analysis, researchers took data from Case Search Information Systems throughout Indonesia. The data is then filtered based on the classification of cases, namely the case of Information and Electronic Transactions that occurred from January 2021 to March 2022. The data that was successfully obtained were 1101 cases. There were still some cases, but the emergence of obstacles, namely case details from Case Search Information Systems, could not be accessed or cases classified. Of the 1101 cases, the data was then cleaned so that it could be analysed using RapidMiner so that there were 449 cases of online fraud.

The researcher categorizes the cases of Information and Electronic Transactions, namely: Fake News, Violent Content, Pornographic Content, Rebellion, Extortion, Defamation, Online Fraud, Online Narcotics Sales, Unpleasant Acts, Online Gambling, Piracy, Online Prostitution, Skimming, and Hate Speech.

A total of 449 cases of Information and Electronic Transactions related to online fraud were investigated more deeply based on cases and evidence. Information obtained from tracing 449 cases of online fraud is: data on gender, data on the area where the case was prosecuted (province), data on evidence, and the mode of online fraud that occurred, namely: online transactions frauds, impersonation, unauthorized access, digital forgery, lottery/online investment fraud, stolen personal data transaction, data forgery, phishing, cyber-extortion, harassment, money laundering, cyber-defamation, e-fencing, hacking, illegal online content, skimming, carding, and piracy.

The data is then tabulated based on gender, the instant messenger used by the perpetrator, the type of cybercrime based on the investigation of each case, and the area (province) of the district court that brought the case. The researcher uses the attribute [DO THE CRIMINAL IN THE CASE USE INSTANT MESSENGER?] as a label to predict other correlated attributes and performs text mining for the [Media] attribute used by the perpetrator and the type of cybercrime committed.

Table 4 shows RapidMiner's text mining results from 499 cyber-fraud cases. Instant Messengers that victims often use are Facebook, WhatsApp, and Instagram.

Table 4. Instant Messenger That Often Used By Cyber Fraud Offender.

Instant Messenger	**Number of Instant Messenger used by Offender**
Facebook	105
Whatsapp	95
Instagram	23
Telegram	6
Facebook Messenger	4
Michat	2
Twitter	1

The text mining results regarding the types of cyber fraud cases that often occur are shown in Table 5. From Table 5, it can be seen that the most common types of online fraud cases are online transactions frauds.

Table 5. Cyber Fraud Case by Categories.

Cyber Fraud	Number of Cyber Fraud on Cases
Online Transactions Frauds	107
Impersonation	25
Unauthorized Access	23
Digital Forgery	20
Lottery	10
Online Investment	10
Buying Stolen Personal Data	9
Data Forgery	6
Phising	6
Cyber-Extortion	5
Harassment	4
Money Laundering	3
Cyber-Defamation	1
E-Fencing	1
Hacking	1
Illegal Online Content	1
Skimming	1

Table 6 shows that the Naïve Bayes and Random Forest Accuracy value is 87.5%, and Recall is 99%. With an accuracy percentage of 87.5%, the prediction results from this Naive Bayes algorithm and Random Forest can be used as data for the element Potential Offender.

Table 6. Performance on Predicting Cyber Fraud Offender Using Confusion Matrix.

Model	Performance			
	Accuracy	Classification Error	Precision	Recall
Naïve Bayes	87.5%	12.5%	86.7%	99%
Decision Tree	76.7	23.3	76.7	100
Random Forest	87.5%	12.5%	86.7%	99%

Based on the comparison between Table 3 and Table 6, it can be concluded that the performance of an adequate algorithm for profiling victims and perpetrators of crimes online is the Naive Bayes algorithm. Thus, profiling is made based on the Naive Bayes algorithm model simulation results.

The analysis using the Naive Bayes Algorithm produces a predictive model, as shown in Table 7, which shows the results of online fraud victim profiles based on the weights for each attribute. This is the model prediction for online fraud victim.

Table 7. Cyber Fraud Victim Profile using Naïve Bayes.

Attribute	Prediction
Age	Between 23 – 28 years old
Number of Instant Messengers Owned	More than 3 Applications
Gender	Female
Education	High School
Major	Management
Occupation	Student/College Student
Region	Special Region of Yogyakarta
Duration of Using the Internet in One Day	More than 8 hours
Gadget	Smartphone (iPhone/Android)
Instant Messenger	Instagram, Facebook, WhatsApp

The information obtained from the Case Tracking Information System and the analysis results using the classification technique using the Naïve Bayes Algorithm theoretically produce online profiling of criminals. Based on the analysis results using data mining classification techniques, the Naïve Bayes Algorithm can create profiling of cyber fraud perpetrators. Table 8 shows the results of the profile of cyber fraud perpetrators in Indonesia.

Table 8. Cyber Fraud Offender Profiling using Naïve Bayes.

Attributes	Prediction
Gender	Male
Region	Capital Special Region of Jakarta
Instant Messenger	Facebook, WhatsApp, Instagram

Based on the analysis using the Naïve Bayes Algorithm, the attributes that can be used as references for the victim's third element of the RAT, Instant Messenger, is obtained. IMs that victims often use are Instagram, WhatsApp, and Facebook. The IM used by the victim is very good in terms of protection against unauthorized access. This result strengthens the theory of Stajano and Wilson[12] that a weak system exists in the element of its users. So in the case of online fraud that occurs in Indonesia, it is the result of attacks by social engineering that attack humans, not applications or devices. The absence of a guardian is not from the application but from the human factor (lack of cyber security knowledge).

Based on the profiling of cyber-fraud perpetrators, data can be obtained about the gender of the perpetrator, the type of online fraud committed, and the scene of the case or the place of trial. The results of online profiling of fraud victims get data on socio-demographic characteristics, such as age, gender, education, domicile, and length of time accessing the internet in a day. Profiling perpetrators and victims of online fraud also reveal information about the devices and applications used to use the internet.

Online profiling of fraud perpetrators is male, whether online fraud using IM or not. In cases recorded in the Case Investigation Information System, 291 male are the perpetrators. In the results of the profiling of the perpetrators, it was also found that the place of occurrence of the case and the domicile of the most perpetrators were in the Province of the Special Capital Region of Jakarta. Based on Part Two, Chapter X, Article 84, Article 85, and Article 86 of Law no. 8 of 1981 concerning the Criminal Procedure Code states that the district court has the authority to hear all cases based on the crime committed (locus delicti) or the residence of the defendant and the residence of most of the witnesses summoned. So it can be

concluded that the perpetrators of crimes are tried in court following the place of occurrence of the case or the domicile of the perpetrators.

Table 9 is a tabulation based on information obtained from the profiling analysis results using the RAT's three elements. The data comes from prediction results for profiling offenders and victims of online fraud and analysis of IM security capabilities.

Table 9. RAT Online Fraud Profiling using Naïve Bayes.

Elements	Attributes	Prediction
Potential Offender	Gender	Male
	Instant Messenger	Facebook, WhatsApp, Instagram
	Region	Capital Special Region of Jakarta
Suitable Target	Gender	Female
	IM installed	More than 3 apps
	Duration of Using the Internet in One Day	More than 8 hours
	Region	Special Region of Yogyakarta
	Instant Messenger	Instagram, WhatsApp, Facebook
Guardianship	Lack of knowledge on cyber-fraud	

Based on Case Search Information Systems, the most common types of online fraud using IM are online transaction fraud, impersonation, illegal access, document forgery, lottery/online investment, trading transaction of customer/personal data (buying stolen personal data), and data forgery.

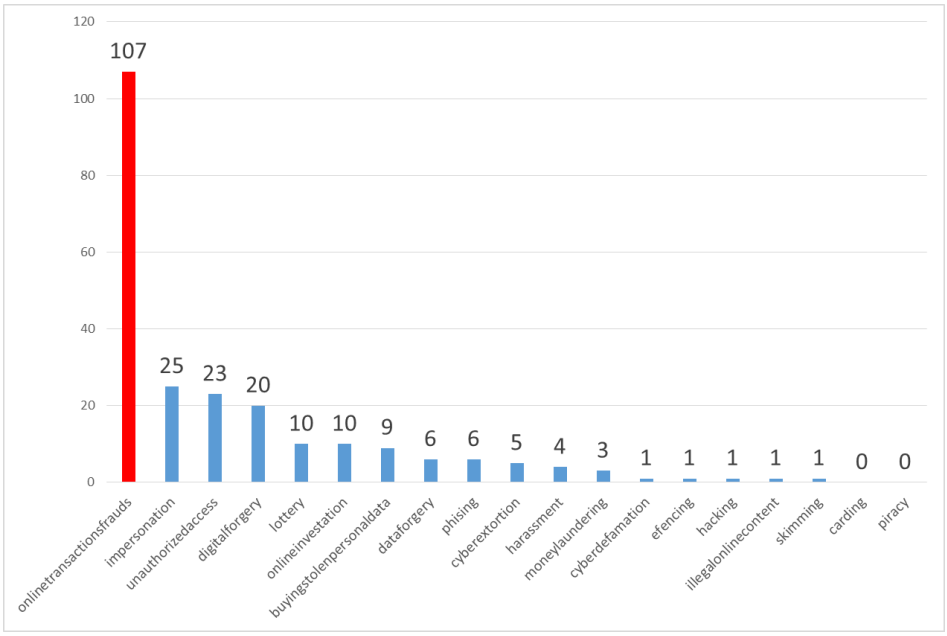

Figure 6. Patterns of online fraud in Indonesia based on structured data.

Figure 6 shows the pattern of online fraud using IM in Indonesia in 2021, with buying and selling transaction fraud getting the highest number, namely 107 cases. There is an allegation that there is a relationship between online shopping triggering victims to enter the perpetrator's social media.

4. Conclusions

Research on the analysis of cyber fraud patterns in Indonesia based on routine activity theory with data mining techniques has successfully concluded that the pattern of online fraud is generally carried out with online transactions frauds, impersonation, unauthorized access, digital forgery, lottery/online investment, buying stolen personal data, and data forgery. These finding was obtained from the results of text mining cases that were tried in the Indonesian District Court from January 2021 to March 2022. Cyber-fraud profiling based on RAT with Naïve Bayes Algorithm yields the following findings: (1) Potential Offender Elements: Male, using Facebook, WhatsApp, and Instagram, and crime scene region in Special Capital Region of Jakarta; (2) Elements Suitable Target: Female, using Instagram, WhatsApp, and Facebook, living in the Special Region of Yogyakarta, spending time on the internet more than 8 hours a day, and have more than three IM applications; and (3) Guardianship: Lack of knowledge about Cyber Fraud.

Acknowledgement

The authors wish to thank the Department of Industry, Ahmad Dahlan University to the giving the avenue to publish this research project.

Funding

The authors received no specific funding for this work.

Author Contribution

Author1 prepared a part of the literature review for the data mining methods, designed the conceptual diagram, wrote the research methodology for the method, implemented the method, and interpreted the result. Author2 created the abstract, prepared the literature review for cybercrime method, wrote the research methodology for the method, implemented the method, and provided conclusions. Author3 did the introduction, prepared a part of the literature review for the routine activity theory, interpreted the results by ranking the alternatives, and wrote the acknowledgment and oversaw the article writing.

Conflict of Interest

In accordance with MJOC policy and my ethical obligation as a researcher, we are reporting that we have no financial and/or business interests in any party that may affect the research reported in the enclosed paper.

References

Agustina, J. R. (2015). Understanding cyber victimization: Digital architectures and the disinhibition effect. *International Journal of Cyber Criminology*, *9*(1), 35–54. https://doi.org/10.5281/zenodo.22239

Ahmad, R., & Thurasamy, R. (2022). A Systematic Literature Review of Routine Activity Theory's Applicability in Cybercrimes. *Journal of Cyber Security and Mobility*, *11*(3), 405–432. https://doi.org/10.13052/jcsm2245-1439.1133

Aimran, N., Rambli, A., Afthanorhan, A., Mahmud, A., Sapri, A., & Aireen, A. (2022). *Prediction of Malaysian Women Divorce Using*. *7*(2), 1067–1081. https://doi.org/10.24191/mjoc.v7i2.17077

Alzubaidi, A. (2021a). Cybercrime Awareness among Saudi Nationals: Dataset. *Data in Brief*, *36*, 106965. https://doi.org/10.1016/j.dib.2021.106965

Alzubaidi, A. (2021b). Measuring the level of cyber-security awareness for cybercrime in Saudi Arabia. *Heliyon*, *7*(1), e06016. https://doi.org/10.1016/j.heliyon.2021.e06016

Bjelajac, Ž., Matijašević, J., & Dimitrijević1, D. (2012). Computer Fraud as a Part of Contemporary Security Challenges. *The Review of International Affairs*, *LXIII*(1147), 5–21.

Bock, K., Shannon, S., Movahedi, Y., & Cukier, M. (2017). Application of Routine Activity Theory to Cyber Intrusion Location and Time. *Proceedings - 2017 13th European Dependable Computing Conference, EDCC 2017*, 139–146. https://doi.org/10.1109/EDCC.2017.24

Choi, K. (2008). Computer Crime Victimization and Integrated Theory: An Empirical Assessment. *International Journal of Cyber Criminology*, *2*(1), 308–333.

Erdoğdu, M., & Koçyiğit, M. (2021). The Correlation between Social Media Use and Cyber Victimization: A Research on Generation Z in Turkey. *Connectist: Istanbul University Journal of Communication Sciences*, 101–125. https://doi.org/10.26650/connectist2021-817567

Goni, O. (2022). Cyber Crime and Its Classification. *Int. J. of Electronics Engineering and Applications*, *May*. https://doi.org/10.30696/IJEEA.X.I.2021.01-17

Hassan, M. M., & Mirza, T. (2018). Customer Profiling and Segmentation in Retail Banks Using Data Mining Techniques. *International Journal of Advanced Research in Computer Science*, *9*(4), 24–29. https://doi.org/10.26483/ijarcs.v9i4.6172

Hawdon, J., Costello, M., Ratliff, T., Hall, L., & Middleton, J. (2017). Conflict Management Styles and Cybervictimization: Extending Routine Activity Theory. *Sociological Spectrum*, *37*(4), 250–266. https://doi.org/10.1080/02732173.2017.1334608

K. Sindhu, K., & B. Meshram, B. (2012). Digital Forensics and Cyber Crime Datamining. *Journal of Information Security*, *03*(03), 196–201. https://doi.org/10.4236/jis.2012.33024

Kigerl, A. (2012). Routine Activity Theory and the Determinants of High Cybercrime Countries. *Social Science Computer Review*, *30*(4), 470–486. https://doi.org/10.1177/0894439311422689

Leukfeldt, E. R. (2014). Phishing for suitable targets in the Netherlands: Routine activity theory and phishing victimization. *Cyberpsychology, Behavior, and Social Networking*,

17(8), 551–555. https://doi.org/10.1089/cyber.2014.0008

Li, X. (2020). Analysis of Criminal Activities Exploiting Social Media: With Special Regards to Criminal Cases of Wechat Fraud in Chinese Jurisdiction. *Journal of Legal Studies*, *26*(40), 19–36. https://doi.org/10.2478/jles-2020-0009

Mahmud, S., Chakraborty, D., Tasnim, L., Tahira, N. J., & Ferdous, M. F. (2020). The Economic Impact of Social Media Fraud and it's Remedies. *International Journal of Machine Learning and Networked Collaborative Engineering*, *4*(1), 30–39. https://doi.org/10.30991/ijmlnce.2020v04i01.004

Mamade, B. K., & Dabala, D. M. (2021). Exploring The Correlation between Cyber Security Awareness, Protection Measures and the State of Victimhood: The Case Study of Ambo University's Academic Staffs. *Journal of Cyber Security and Mobility*, *10*(4), 699–724. https://doi.org/10.13052/jcsm2245-1439.1044

Marshal, A. M. (2009). *Digital Forensics Digital Evidence in Criminal Investigations* (1st Ed). Wiley-Blackwell.

Michael, G. (2020). Knowledge Based System for Predicting Cyber Crime Patterns Using Data Mining. *Journal Of Critical Reviews*, *7*(10), 2043–2053.

Ngo, F., & Paternoster, R. (2011). Cybercrime Victimization: An examination of Individual and Situational level factors. *International Journal of Cyber Criminology*, *5*(1), 773–793.

Palaniappan, S., Mustapha, A., Foozy, C. F. M., & Atan, R. (2017). Customer profiling using classification approach for bank telemarketing. *International Journal on Informatics Visualization*, *1*(4–2), 214–217. https://doi.org/10.30630/joiv.1.4-2.68

Reep-van den Bergh, C. M. M., & Junger, M. (2018). Victims of cybercrime in Europe: a review of victim surveys. *Crime Science*, *7*(1). https://doi.org/10.1186/s40163-018-0079-3

Ritonga, A. S., & Muhandhis, I. (2021). Teknik Data Mining Untuk Mengklasifikasikan Data Ulasan Destinasi Wisata Menggunakan Reduksi Data Principal Component Analysis (Pca). *Edutic - Scientific Journal of Informatics Education*, *7*(2). https://doi.org/10.21107/edutic.v7i2.9247

Saroha, R. (2014). Profiling a Cyber Criminal. *International Journal of Information and Computing Technology*, *4*(3), 253–258.

Schreck, C. J. (2017). Routine Activity Theory. *Preventing Crime and Violence*, 67–72. https://doi.org/10.1007/978-3-319-44124-5_7

Sebastian, S. R., Babu, B. P., & Sebastian, S. R. (2023). *Are we cyber aware ? A cross sectional study on the prevailing cyber practices among adults from Thiruvalla , Kerala*. *10*(1), 235–239. https://doi.org/10.18203/2394-6040.ijcmph20223550

Shahira Pisal, N., Abdul-Rahman, S., Hanafiah, M., & Kamarudin, S. I. (2022). Prediction of Life Expectancy for Asian Population Using Machine Learning Algorithms. *Malaysian Journal of Computing*, *7*(2), 1150–1161. https://doi.org/10.24191/mjoc.v7i2.18218

Shrivastava, R., & Jain, R. (2021). Impact of Cyber Crime on Youth in Lockdown. *Legal Research Development: An International Refereed e-Journal, Vol. 6*(Issue-I), 15–20. https://doi.org/https://doi.org/10.53724/lrd/v6n1.04

Sianturi, C. M., Pasaribu, V. A. R., Pasaribu, R. M., & Simanjuntak, J. (2022). the Impact of Social Media Marketing on Purchase Intention. *SULTANIST: Jurnal Manajemen Dan Keuangan*, *10*(1), 60–68. https://doi.org/10.37403/sultanist.v10i1.425

Singh, N. P. (2007). Online Frauds in Banks with Phishing. *Journal of Internet Banking and Commerce*, *12*(2), 1–28. http://eprints.utm.my/8136/

Stajano, F., & Wilson, P. (2011). Understanding scam victims. *Communications of the ACM*, *54*(3), 70–75. https://doi.org/10.1145/1897852.1897872

Sumirat, J. R. (2021). Policing on Preventing Cyber Fraud in Indonesia. *University of York Social Policy Social Work*, *September 2020*, 1–69. https://doi.org/10.13140/RG.2.2.14771.55844

Sunardi, Fadlil, A., & Kusuma, N. M. P. (2022). Implementasi Data Mining dengan Algoritma Naïve Bayes untuk Profiling Korban Penipuan Online di Indonesia. *Jurnal Media Informatika Budidarma*, *6*, 1562–1572. https://doi.org/10.30865/mib.v6i3.3999

Sunardi, Fadlil, A., & Kusuma, N. M. P. (2023). Comparing Data Mining Classification for Online Fraud Victim Profile in Indonesia. *Intensif*, *7*(1), 1–17. https://doi.org/10.29407/intensif.v7i1.18283

Tompsett, E. I. B. C., Marshall, A. M., & Semmens, N. C. (2005). Cyberprofiling: Offender profiling and geographic profiling of crime on the internet. *Workshop of the 1st International Conference on Security and Privacy for Emerging Areas in Communication Networks*, *2005*, *2005*, 22–25. https://doi.org/10.1109/SECCMW.2005.1588290

Yar, M. (2005). The Novelty of 'Cybercrime': An Assessment in Light of Routine Activity Theory. *European Journal of Criminology*, *2*(4), 407–427. https://doi.org/10.1177/147737080556056