Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12422 p-ISSN : 2541-044X

Crime of theft prediction using Machine Learning K-Nearest Neighbour Algorithm at Polresta Bandar Lampung

Febry Hermawan[1)*], Jarot Prianggono[2)]
[1,2)]Sekolah Tinggi Ilmu Kepolisian PTIK ,Jl. Tirtayasa No.6, RW.4, Melawai, Kec. Kby. Baru, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12160
[1)]febryhermawan333@yahoo.com,[2)]komputerstik@gmail.com

Submitted: May 13, 2023 | Accepted : May 23, 2023 | Published : Jul 1, 2023

Abstract: The era of the industrial revolution 4.0 is a time where cyber and physical technology collaborate. This study aims to predict the types of theft crimes that occur in the Bandar Lampung Police area with the K-Nearest Neighbor algorithm, evaluate the prediction results and profiling the prediction results carried out by Bandar Lampung Police investigators in efforts to prevent and handle criminal acts of theft in the jurisdiction of the Bandar Lampung Police Lampung. The approach was carried out using the quantitative method of the K-Nearest Neighbor algorithm using the Rapidminer application by utilizing 1671 police report data from the Bandar Lampung Police and a questionnaire survey method conducted on 49 police investigators from the Bandar Lampung Police. Data collection techniques are carried out in a valid and reliable manner as a support for predictive validity. Based on the results of the classification and questionnaire, it was found that the majority of victims of the crime of theft were adult men who did not have a job and lived in urban areas. It was found that the majority of thefts occurred in parking lots in urban areas on Monday morning where the perpetrators used tools and targeted moving objects by tampering with locks which caused losses of around 10-50 million rupiah. This type of theft is theft by weighting (CURAT) which applies to Article 363 of the Criminal Code. The prediction results show that the neighboring value (K) and the distribution ratio of training and testing data are K=3 and 7:3, respectively. Predictions using K values and data sharing ratios show a high level of accuracy, namely 99. 20%. The results of the questionnaire show results that are in line with the results of the classification with an accuracy rate of the actual data of 75. 7122%. So by increasing the understanding skills of Bandar Lampung Police investigators using technology to predict the crime of theft, the number of theft crimes can be reduced.

Keywords: k-NN Algorithm, Era 4.0, Victims of theft, Prediction

INTRODUCTION

The era of industrial revolution 4.0 is a time in which cyber and physical technology collaborate. The presence of this era marks the continuity of change in line with the human ability to produce innovation. Over time, each period of time has always produced extraordinary

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12422 p-ISSN : 2541-044X

innovations, with a focus on meeting needs and improving the quality of life. Other features of the industrial revolution 4.0 era include the Internet of Things (IoT), cloud-based computing, and cognitive computing, or what we call Machine Learning (ML). In the era of the industrial revolution 4.0, technological developments went hand in hand with innovations created by humans (Prasetyo, 2020).

However, the Industrial Revolution 4.0 brought more fundamental changes, with a far-reaching impact on the way humans think, live and interact with one another. This era will not only change various aspects of technology, but will also affect other fields such as economics, social and politics (Prasetyo, 2020). Human activities in various sectors will experience significant disruption due to this change, requiring new adaptations and innovations to deal with it. In facing this challenge, new opportunities arise to optimize potential and overcome obstacles that may arise.

Theft is a form of crime that often occurs. The motivations behind acts of theft can vary, from the desire to give gifts to someone special, as is often done by teenagers, to meeting basic daily needs (Jamilah et al., 2020). In some cases, theft may also be triggered by other factors, such as economic instability, social pressure, or opportunities that arise. Although the reasons behind acts of theft may vary, these actions are still considered unlawful acts that harm other parties and undermine public security and order. Therefore, prevention efforts, law enforcement, and awareness of the consequences of acts of theft remain important in maintaining security and justice in society.

Table 1. 2019-2022 Report on Criminal Acts of Theft

No	Type of Crime	Year				Total
		2019	2020	2021	2022	
1.	Cubis	102	110	108	105	425
2.	Curas	62	67	51	34	215
3.	Curat	155	345	266	265	1031
Total		319	522	426	404	1671

Source: Data Min Ops Criminal Investigation Police Bandar Lampung 2019-2022

The list presents the number of theft crimes in Bandar Lampung between 2019 and 2022 which is quite high, such as 1031 cases of weighted theft (CURAT). This is related to the outbreak of the COVID-19 pandemic which began in 2019 and spread to Indonesia. Legal countermeasures, which are based on accurate data using technology, such as the use of artificial intelligence (AI) (Rachmadie, 2020).

This research uses data mining which comes from computer data and identification records of theft cases. The data is then processed using the k-NN algorithm. The k-NN algorithm is an information grouping technique that utilizes training information that is closest to the data being searched for. This algorithm is included in the category of supervised learning, where new data can be classified based on the majority of k-NN classes. The most commonly used categories as grouping categories. This method allows researchers to analyze theft data effectively, identify patterns, and classify data using k-NN as the basis for the algorithm (Hakim et al., 2019). Thus, Data Mining techniques and the k-NN algorithm can be used as an efficient approach in analyzing theft data to retrieve valuable information and support crime prevention and control efforts.

Previous research by (Umoh et al., 2021) found that k-NN had the highest accuracy in predicting crime with 0.967597, followed by the SVM algorithm with an accuracy rate of 0.965828 and RF with an accuracy rate of 0. 964973. Research (Umair et al., 2020) uses data on crime in Pakistan, resulting in the conclusion that the accuracy of k-NN can reach a maximum of 92%, while the Random Forest algorithm can only reach 62%. The results showed that the higher the value of K, the better the k-NN algorithm, where for K values of 3, 5, 7, and 9 the k-NN algorithm has an accuracy rate of 89%, 93%, 93%, and 94% .

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12422 p-ISSN : 2541-044X

LITERATURE REVIEWS

Police Science

In this research, the basic concept of policing science is used to utilize the field of technology science in handling social problems in the jurisdiction of Polresta Bandar Lampung. The main duties of the Republic of Indonesia National Police in Law No. 2 of 20002 Article 13 are as follows: Maintain public order and security, uphold the law, provide protection, protection and service to the community. ", the description of the Police's duties is explained again in Article 14 of the Indonesian Police Law (Polri humas, 2023). The main task of the police in maintaining public order and security requires the professionalism of its personnel. Therefore, organizations need to ensure that personnel have an in-depth understanding of various scientific fields, including the ability to make maximum use of technology to support police duties. The use of police technology and a crime prevention approach in this study is consistent with the multidisciplinary nature of police science. Through this approach, it is hoped that a holistic understanding and effective solutions will be obtained in dealing with the crime problems faced in the region. By combining the basic concepts of police science and technology, this research aims to present a comprehensive and innovative approach in efforts to deal with crime and maintain public safety.

Knowledge Discovery In Database

Knowledge Discovery in Database is a non-trivial process that is used in pattern recognition or relationships that are valid, new, have potential benefits, and can ultimately be understood in data sets in making important decisions (Brodley et al., 1999). More information can reduce errors (errors) and increase the use of the model as a benchmark. Data mining is an activity that combines statistics, mathematics, artificial intelligence and machine learning. The goal is to extract and identify a variety of existing data to generate new information that is useful for future results in large databases (Handoko, 2016).

Data Mining

Data mining is a field that involves extracting valuable information from large data sets. Data mining is divided into various sets based on the tasks to be performed, namely Description, Estimation, Prediction, Classification, Clustering, Association. In order to achieve these goals, data mining can be grouped into several sets based on the tasks to be achieved. First, the Description task aims to provide a brief yet informative description of the characteristics and patterns present in the data. Next, the Estimate task aims to calculate or estimate unknown values in the data. Prediction is the task of building a model that can be used to predict future values or events based on existing data. Classification is the task of grouping data into predetermined categories or classes based on existing attributes. Clustering, on the other hand, aims to group data into similar or interrelated groups based on certain characteristics. Finally, the Association task aims to find relationships or correlations between items or attributes in the data. By dividing data mining into these categories, researchers and practitioners can select and apply techniques appropriate to their analytical goals, thereby enabling them to gain valuable insights from the existing data.

Theft Crime

The crime of theft is regulated in the Indonesian Criminal Code (KUHP). Theft is defined as the act of taking or taking possession of another person's property without the rights or permission of the owner, with the intent to permanently control the item and to gain profit for oneself or another person. Book II of the Criminal Code has been perfectly formulated, meaning that the formulation contains complete elements, both objective and subjective elements. Objective elements can be elements of material actions, elements of objects or goods, elements of conditions accompanying object objects, elements of efforts to carry out prohibited actions, elements of constitutive consequences. Subjective elements can be; elements of error, elements against the law (Chazawi, 2002). In the Criminal Code, the crime of theft is regulated in Articles 362 to Article 368. Several important elements that must be fulfilled to constitute the crime of

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12422 p-ISSN : 2541-044X

theft include the act of taking or removing goods, the presence of an element of ownership by another person, and the presence of malicious intent to gain illegal profits. In addition, the Criminal Code also regulates variations of criminal acts of theft, such as aggravated theft, aggravated theft at home, aggravated theft or violence, and aggravated theft resulting in death or serious injury.

METHODS

In this study, a quantitative research approach was used, namely research using the number method to describe the observation of an object or variable, numbers being part of the measurement. The method used is the k-NN algorithm method, which is a data mining method used for classification and prediction of a data set so that it can be used for decision making by related parties. In this study, a quantitative research approach was carried out using the number method to describe the observation of an object or variable. This method is based on the use of numbers as part of the measurements taken. In this context, the method used is the k-NN (k-Nearest Neighbors) algorithm. This method is one of the methods used in data mining to classify and predict existing data sets. Through the use of this method, the information obtained from the data can be used for decision making by related parties. By utilizing the k-NN algorithm, this study aims to provide deeper insight and understanding of the object or variable being studied, and to be able to contribute to decision making related to the research context.

K-NN Method Flow

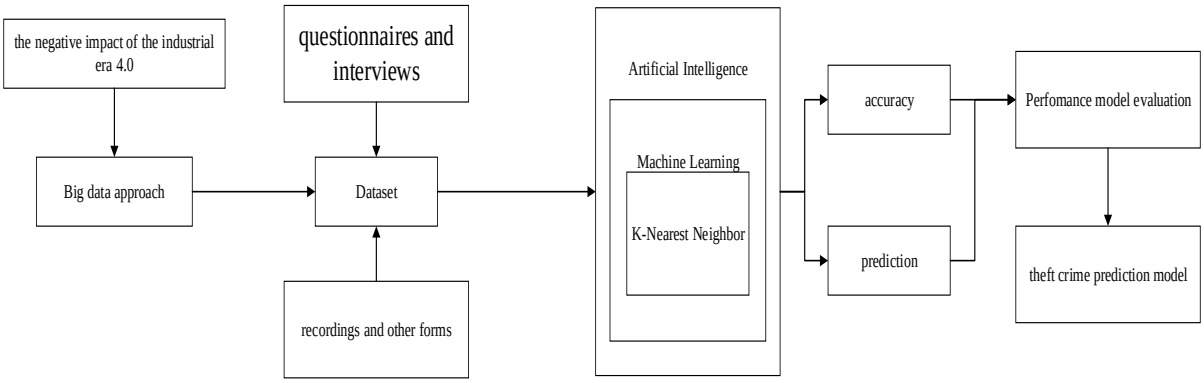


Figure 1. The flow of the k-NN algorithm method in research

RESULTS AND DISCUSSIONS

In making predictions using the k-NN algorithm with Rapidminer, the first thing that must be done is data cleansing of the data that has been collected through reports of victims of criminal acts of theft in the Bandar Lampung Police area. After the data is collected, the data is then cleaned through the Data Cleansing process where the data that has been collected as many as 1671 reports are adjusted so that they are easy to use in the data processing, the parts that must be simplified to facilitate the processing of the data such as the writing font, the size of the letters, categorization of numeric data, and also clears data from punctuation. After going through the data cleansing process, the data sets that are ready for use will then be imported into the local repository of the Rapidminer application. After being imported, the data is then adjusted according to the type/type of each variable for both polynominal (many types of data) and binomial (only 2 types of data). Then the variable you want to predict changes its role to a label. Then the data is ready to be imported into the local repository. After the data has been successfully imported, the next thing that must be done is to build a model design in the Rapidminer application.

In the process of building a design model, the data retrieve operator plays an important role in retrieving data from the local repository to be processed according to the model design that has been built. This operator acts as a conduit connecting the data with the next steps in the analysis

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12422 p-ISSN : 2541-044X

process. After the data has been successfully retrieved, the next step is the data split operator which has a significant role. This operator aims to divide data into smaller sets, which will then be subject to model building using the k-NN algorithm in the RapidMiner application. In the process of classifying data, dividing data into smaller subsets allows for more focused and efficient analysis. By separating the data into subsets, each subset can be processed separately to better understand its characteristics. This enables the development of more accurate and effective models in the k-NN algorithm. In the RapidMiner application, the use of the data split operator allows the researcher to choose how the data will be divided into the relevant subsets. This can be done based on certain criteria such as the distribution of data based on attributes, the desired percentage of data for each subset, or other methods of dividing data according to research needs. By utilizing the data retrieve operator and data split operator in the RapidMiner application, researchers can optimize the data processing process, build the right model, and produce more accurate results using the k-NN algorithm.

As the name implies, the k-NN algorithm uses the constant k as a neighbor constant, a unit of dimension in the distance between data. This constant is calculated to compare the data that will be predicted by the model design earlier with the training data which is the nearest neighbor data from the data to be predicted. After going through the calculations with the algorithm above, the data results will be processed by the performance operator to find out the percentage accuracy of the results compared to the predicted statistical results.

In this study, the classification used to predict the type of theft crime that will occur at a certain time is theft with violence (CURAS), ordinary theft (CUBIS) or aggravated theft (CURAT). In order to evaluate the statistical performance of a classification model, the data set must be labeled i.e. it must have an attribute with the role of the label. The label attribute stores the actual observed value of the value that would be predicted by the classification model being discussed. In the model made above, the per and exa axes are connected to the right side of the process area to display the evaluation of the accuracy percentage and also display the prediction results statistics.

In this study, the authors have designed a Machine Learning (ML) architecture that uses the k-NN Classifier algorithm. This architecture was developed with the aim of predicting theft crimes that occurred in the jurisdiction of the Bandar Lampung Police. The image shown below visually illustrates the ML architecture that has been designed by the author. This architecture includes various important components, including the data preprocessing stage, the formation of the k-NN Classifier model, as well as the model evaluation and validation stage. By utilizing this architecture, it is hoped that this research can provide accurate and useful prediction results in efforts to prevent and handle theft crimes at the Bandar Lampung Police.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12422 p-ISSN : 2541-044X

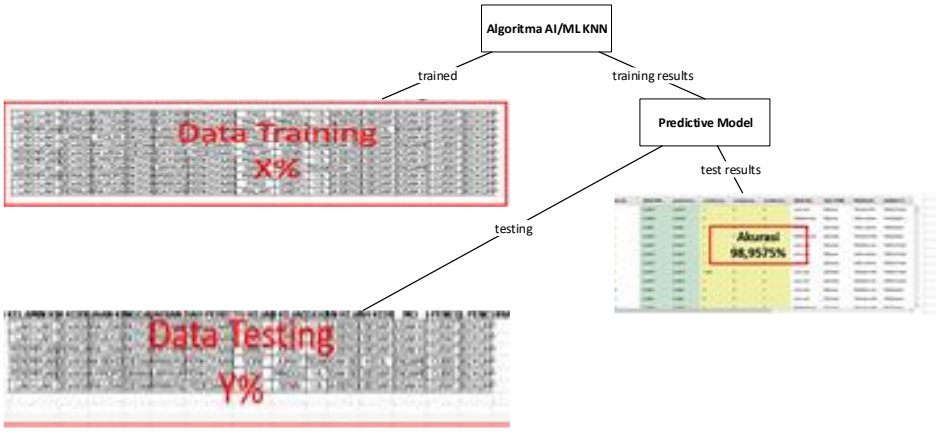


Figure 2. Architecture of the k-NN Machine Learning Method in Research

The X and Y variables in this study have been varied according to the existing needs. The results of changing these variables are then displayed in the following table::

Table 2. Predicted Value of Theft Crimes in 2019-2022

K	Ratio Training: Testing 6:4	Ratio Training: Testing 7:3	Ratio Training: Testing 8:2	Ratio Training: Testing 9:1	
1	98,35	98,4	98,5	99,4	
3	98,65	99,2	99,1	99,4	
5	98,8	99	99,1	99,4	
7	98,65	99	98,8	99,4	
9	98,8	99	98,8	99,4	
Average	98,65	98,92	98,86	99,4	98,9575

Based on the calculation results, it was found that the model that had been developed was able to predict the crime of theft at the Bandar Lampung Police with an average accuracy rate of 98.9575%. The highest predictive results were obtained from a combination of training and testing data ratios of 90% for training and 10% for testing, with an accuracy rate of 99.40%. These results indicate that the model that has been designed has a very good ability to predict theft crimes in the Bandar Lampung Police area, with a very high level of accuracy. This gives an indication that the model can be used as an effective tool in supporting efforts to prevent and handle theft crimes in the region.

Predictions

Umair et al. (2020), the accuracy of predictions with the k-Nearest Neighbor algorithm will increase at odd K values and will reach its maximum value at k=9. Based on this research, the researcher took the value K = 1, 3, 5, 7, 9, each of which has results according to the distribution of the data.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12422 p-ISSN : 2541-044X

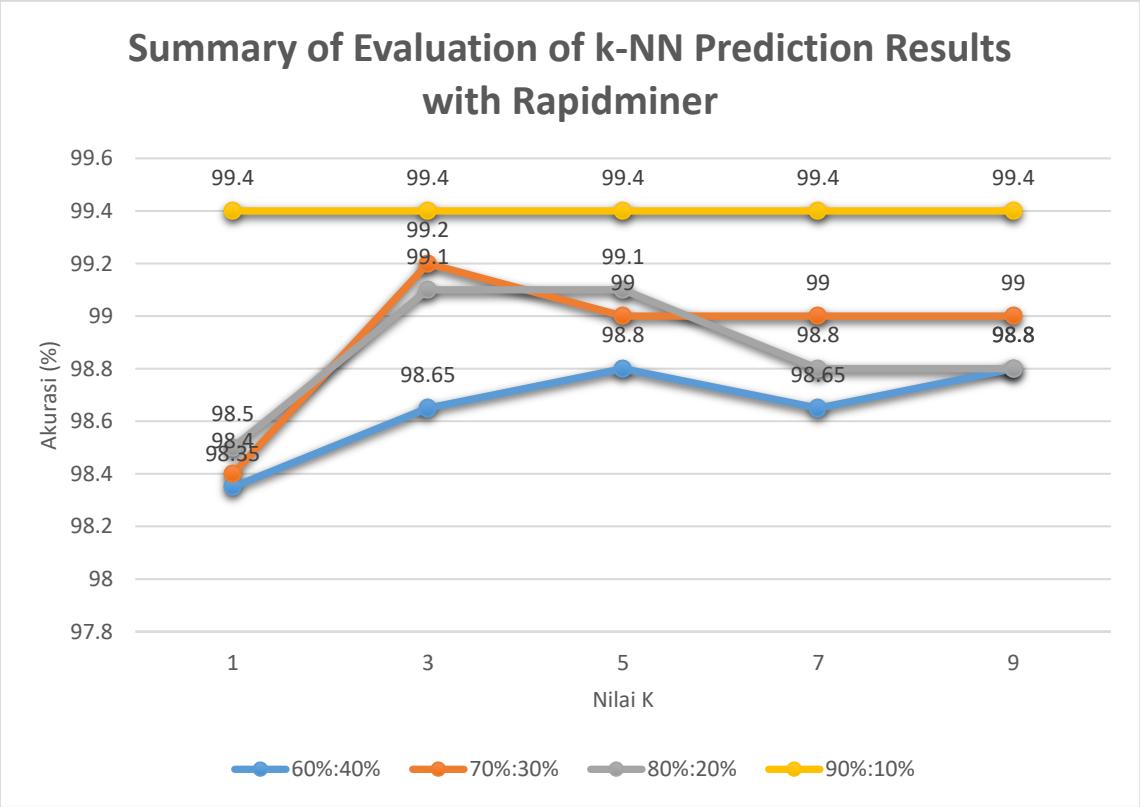


Figure 3. K-NN Prediction Results
Source: Results of Data Processing

accuracy: 99.40%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	103	1	0	99.04%
pred. CURAS	0	20	0	100.00%
pred. CUBIS	0	0	42	100.00%
class recall	100.00%	95.24%	100.00%	

Figure 4. Prediction results of k-NN with K = 1 and 10% Testing Data

accuracy: 99.40%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	103	1	0	99.04%
pred. CURAS	0	20	0	100.00%
pred. CUBIS	0	0	42	100.00%
class recall	100.00%	95.24%	100.00%	

Figure 5. Prediction results of k-NN with K = 3 and 10% of data testing

accuracy: 99.40%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	103	1	0	99.04%
pred. CURAS	0	20	0	100.00%
pred. CUBIS	0	0	42	100.00%
class recall	100.00%	95.24%	100.00%	

Figure 6. Prediction results of k-NN with K = 5 and 10% Testing Data

accuracy: 99.40%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	103	1	0	99.04%
pred. CURAS	0	20	0	100.00%
pred. CUBIS	0	0	42	100.00%
class recall	100.00%	95.24%	100.00%	

Figure 7. Prediction results of k-NN with K = 7 and 10% of data testing

accuracy: 99.40%

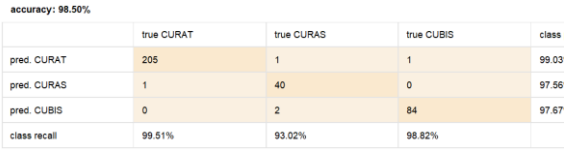
	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	103	1	0	99.04%
pred. CURAS	0	20	0	100.00%
pred. CUBIS	0	0	42	100.00%
class recall	100.00%	95.24%	100.00%	

Figure 8. Prediction results of k-NN with K = 9 and 10% of data testing

From the evaluation of the prediction results with a data testing constant of 10%, it can be seen that all training and testing data sets have the same comparison, namely an accuracy rate of 99.40%. According to these data, the k-NN algorithm shows a very good ability to predict a model design that represents a study because the resulting accuracy rate is very high and almost perfect. That is, with varying neighborhood constants with testing data of 10% it can represent the model very well.

*name of corresponding author

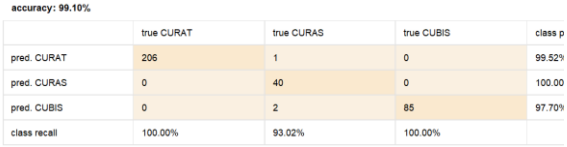
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12422 p-ISSN : 2541-044X



accuracy: 98.50%

	true CURAT	true CURAS	true CUBIS	class
pred. CURAT	205	1	1	99.03
pred. CURAS	1	40	0	97.56
pred. CUBIS	0	2	84	97.67
class recall	99.51%	93.02%	98.82%	

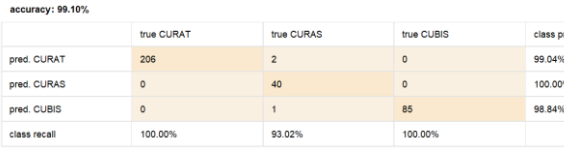
Figure 9. Prediction results of k-NN with K = 1 and 20% Testing Data



accuracy: 99.10%

	true CURAT	true CURAS	true CUBIS	class p
pred. CURAT	206	1	0	99.52%
pred. CURAS	0	40	0	100.00
pred. CUBIS	0	2	85	97.70%
class recall	100.00%	93.02%	100.00%	

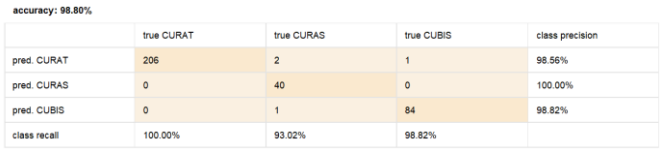
Figure 10. Prediction results for k-NN with K = 3 and 20% of the testing data



accuracy: 99.10%

	true CURAT	true CURAS	true CUBIS	class p
pred. CURAT	206	2	0	99.04%
pred. CURAS	0	40	0	100.00
pred. CUBIS	0	1	85	98.84%
class recall	100.00%	93.02%	100.00%	

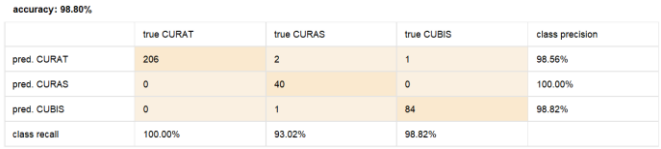
Figure 11. Prediction results of k-NN with K = 5 and 20% Testing Data



accuracy: 98.80%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	206	2	1	98.56%
pred. CURAS	0	40	0	100.00%
pred. CUBIS	0	1	84	98.82%
class recall	100.00%	93.02%	98.82%	

Figure 12. Prediction results for k-NN with K = 7 and 20% of data testing

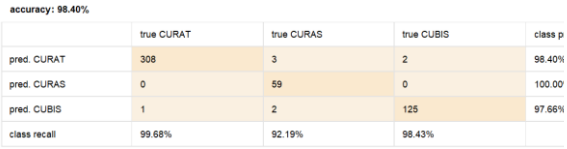


accuracy: 98.80%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	206	2	1	98.56%
pred. CURAS	0	40	0	100.00%
pred. CUBIS	0	1	84	98.82%
class recall	100.00%	93.02%	98.82%	

Figure 13. Prediction results for k-NN with K = 9 and 20% of data testing

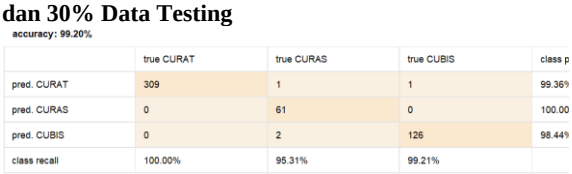
From the evaluation of the prediction results with a data testing constant of 20%, it can be seen that the highest level of accuracy from the training and testing data sets is obtained with the neighboring constant values equal to 3 and 5 with an accuracy value of 99.10%. According to these data, the k-NN algorithm shows a very good ability to predict a model design that represents a study because the resulting accuracy rate is very high. That is, with the neighboring constants equal to 3 and 5 with a testing data of 20%, it can represent the model very well.



accuracy: 98.40%

	true CURAT	true CURAS	true CUBIS	class p
pred. CURAT	308	3	2	98.40%
pred. CURAS	0	59	0	100.00
pred. CUBIS	1	2	125	97.66%
class recall	99.68%	92.19%	98.43%	

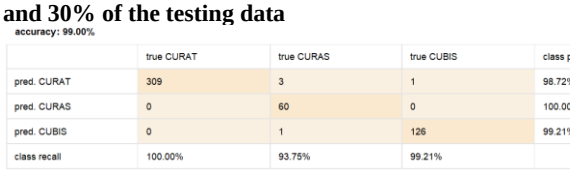
Gambar 14. Hasil Prediksi k-NN dengan K = 1 dan 30% Data Testing



accuracy: 99.20%

	true CURAT	true CURAS	true CUBIS	class p
pred. CURAT	309	1	1	99.36%
pred. CURAS	0	61	0	100.00
pred. CUBIS	0	2	126	98.44%
class recall	100.00%	95.31%	99.21%	

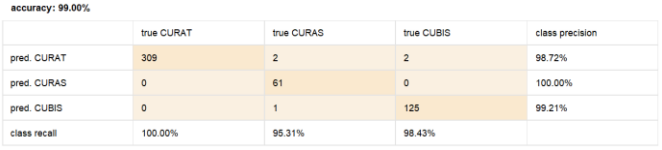
Figure 15. Prediction results for k-NN with K = 3 and 30% of the testing data



accuracy: 99.00%

	true CURAT	true CURAS	true CUBIS	class
pred. CURAT	309	3	1	98.72%
pred. CURAS	0	60	0	100.00
pred. CUBIS	0	1	126	99.21%
class recall	100.00%	93.75%	99.21%	

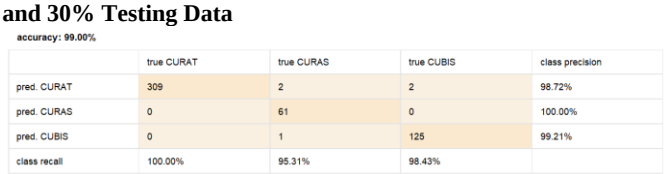
Figure 16. Prediction results of k-NN with K = 5 and 30% Testing Data



accuracy: 99.00%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	309	2	2	98.72%
pred. CURAS	0	61	0	100.00%
pred. CUBIS	0	1	125	99.21%
class recall	100.00%	95.31%	98.43%	

Figure 17. Prediction results of k-NN with K = 7 and 30% Testing Data



accuracy: 99.00%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	309	2	2	98.72%
pred. CURAS	0	61	0	100.00%
pred. CUBIS	0	1	125	99.21%
class recall	100.00%	95.31%	98.43%	

Figure 18. Prediction results for k-NN with K = 9 and 30% of the testing data

From the evaluation of the prediction results with a data testing constant of 30%, it can be seen that the highest level of accuracy from the training and testing data set is obtained with a neighbor constant value equal to 3 with an accuracy value of 99.20%. According to these data, the k-NN algorithm shows a very good ability to predict a model design that represents a study because the resulting accuracy rate is very high. That is, with a neighbor constant equal to 3 with a testing data of 30%, it can represent the model very well.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12543 p-ISSN : 2541-044X

accuracy: 98.35%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	410	5	3	98.09%
pred. CURAS	0	80	0	100.00%
pred. CUBIS	2	1	167	98.24%
class recall	99.51%	93.02%	98.24%	

Figure 19. Prediction results for k-NN with K = 1 and 40% of the testing data

accuracy: 98.65%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	410	4	2	98.56%
pred. CURAS	0	81	0	100.00%
pred. CUBIS	2	1	168	98.25%
class recall	99.51%	94.19%	98.82%	

Figure 20. Prediction results of k-NN with K = 3 and 40% Testing Data

accuracy: 98.80%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	411	4	2	98.56%
pred. CURAS	0	81	0	100.00%
pred. CUBIS	1	1	168	98.82%
class recall	99.76%	94.19%	98.82%	

Figure 21. Prediction results of k-NN with K = 5 and 40% Testing Data

accuracy: 98.65%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	411	4	3	98.33%
pred. CURAS	0	81	0	100.00%
pred. CUBIS	1	1	167	98.82%
class recall	99.76%	94.19%	98.24%	

Figure 22. Prediction results of k-NN with K = 7 and 40% Testing Data

accuracy: 98.80%

	true CURAT	true CURAS	true CUBIS	class precision
pred. CURAT	411	3	3	98.56%
pred. CURAS	0	82	0	100.00%
pred. CUBIS	1	1	167	98.82%
class recall	99.76%	95.35%	98.24%	

Figure 23. Prediction results of k-NN with K = 9 and 40% Testing Data

From the evaluation of the prediction results with a data testing constant of 40%, it can be seen that the highest level of accuracy from the training and testing data sets is obtained with the neighboring constant values equal to 5 and 9 with an accuracy value of 98.80%. The k-NN algorithm shows a good ability to predict a model design that represents a study because the resulting accuracy rate is very high. That is, with the neighboring constants equal to 5 and 9 with data testing of 40%, it can represent the model well.

From all evaluations of prediction results, it can be seen that the comparison of training and testing data sets that has the highest level of accuracy is a ratio of 9:1 with an accuracy of 99.40%, followed by a comparison of 7:3 with an accuracy of 99.20%, then a comparison of 8:2 with an accuracy of 99.10% and finally a comparison of 6:4 with an accuracy of 98.80%. Therefore, the researcher chose a 7:3 ratio with a value of K = 3 with the consideration that the amount of testing data is not too small so that the predicted data results can be more accurate when compared to the original dataset. These results even exceed the results of previous research conducted by (Umair et al., 2020) where for K = 3 the results of this study showed an accuracy rate of 99.20% while the research conducted by Umair only showed 89% results.

Performance Test

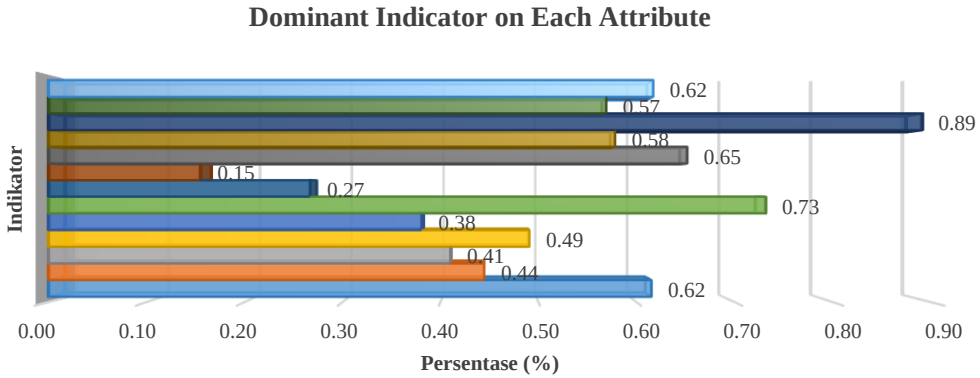


Figure 24. Dominant Indicators of the Profiling Stage (Source: Data Processing Results)

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12543 p-ISSN : 2541-044X

From the picture above, it can be concluded that according to the data on reports of victims of criminal acts of theft that occurred in the jurisdiction of the Bandar Lampung Police, the majority of victims of criminal acts of theft are male in the adult age range who do not have a job, while the rest are victims who work or are semi-employed. -Work. The crime of theft is common in urban areas. The majority of theft cases occur in parking areas with a large percentage. This crime usually occurs in the morning, although it is possible that it will occur during the day or night. Most of these theft crimes usually occur on Mondays. In carrying out the action, the perpetrators of the crime of theft do not hesitate to use tools to help smooth the action, both sharp weapons and blunt weapons, by targeting moving objects and causing huge losses for the victims. The mode of operation used by the perpetrator was to damage the key to steal the target's goods which fall into the category of weighted theft (CURAT) which is included in article 363 of the Criminal Code.

Age

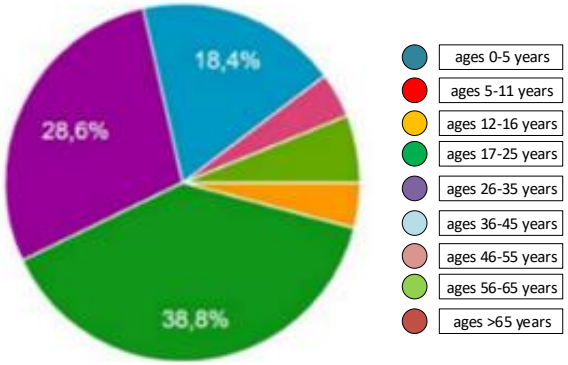


Figure 25. Age of Victims of the Crime of Theft
(Source: Results of Data Processing

The results of research conducted by the author in the jurisdiction of the Bandar Lamping Police showed that there was a dominance of theft victims by certain age groups. Based on the data obtained, it was found that the late adolescent age group, which ranges from 17 to 25 years, is the most vulnerable group to become victims of theft with a percentage of 38.8%. Furthermore, the early adult age group, ranging from 26 to 35 years, occupied the next position with a percentage of 28.6%. Meanwhile, the late adult age group, which ranges from 36 to 45 years, occupies the third position with a percentage of 18.4%. These findings indicate that the late adolescent age group has a higher risk of becoming a victim of theft crime in that area. This may be due to factors such as a lack of supervision or caution by young people in guarding valuables, as well as social and environmental factors that influence crime rates in the area. This research provides a better understanding of the characteristics of victims of theft, so that it can be used as a basis for formulating more effective prevention and protection strategies, especially for the age group that is more vulnerable to becoming victims of this crime.

Genesis Zone

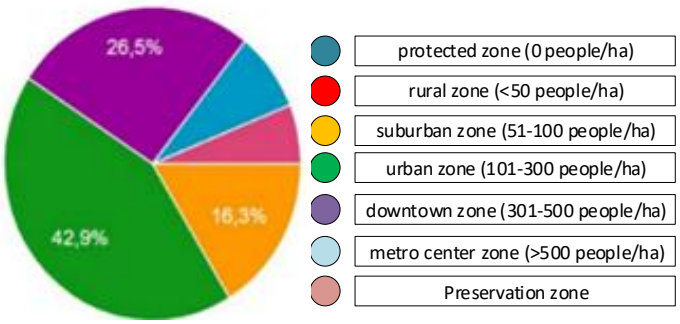


Figure 26. Zone of theft crime
(Source: Data Processing Results)

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12543 p-ISSN : 2541-044X

Based on the analysis conducted, criminal acts of theft often occur in urban zones with a population density of between 101 and 300 residents per hectare. The results showed that the urban zone is the most vulnerable area to theft, with a percentage of 42.9%. It is followed by the downtown zone, which has a population density of between 301 and 500 residents per hectare, at 26.5%. Meanwhile, the suburban zone, which has a population density of between 51 and 100 residents per hectare, occupies the third position with a percentage of 16.3%.

This finding indicates that the crime rate of theft tends to increase in line with an increase in population density in an area. Urban zones, which generally have dense and growing activities, are the main targets for criminals. Factors such as accessibility, crowds and wealth that are concentrated in urban zones can be decisive factors in increasing crime rates in these areas. The results of this study can provide an important basis for designing crime prevention strategies, especially in urban zones that have a higher risk of theft. Efforts to strengthen security and supervision in the zone need to be increased in order to protect the community and reduce the number of theft crimes that occur.

Article Criminal Act of Theft

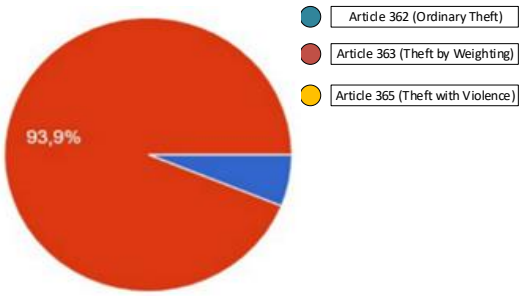


Figure 27. Article of the Crime of Theft
(Source: Data Processing Results)

The results of the research conducted show that most of the criminal acts of theft that occur violate the provisions of the Criminal Code, in particular Article 363 which regulates weighted theft. The percentage of violations of Article 363 reached 93.9% of the total cases studied. This shows that aggravated theft is the most dominant form of theft crime and often occurs in the area studied. Apart from violations of Article 363, there were also a number of cases where violations of Article 362 concerning ordinary theft and Article 365 concerning theft with violence. Even though the percentage is lower than violations of Article 363, it remains an important concern in efforts to deal with the crime of theft as a whole.

This finding illustrates that weighted theft is the most common type of crime and contributes significantly to the overall number of theft cases. Therefore, law enforcement and prevention efforts must be focused on disclosing and handling cases of theft with weighting in order to reduce crime rates and maintain public safety.

Scene

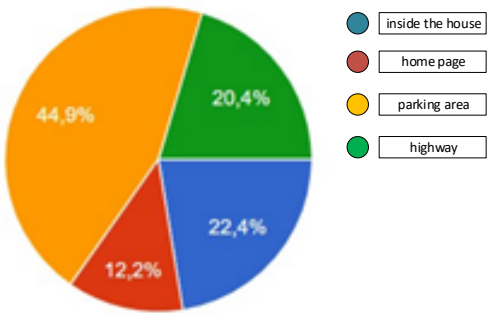


Figure 28. Place of theft crime

(Source: Data Processing Results)

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.12543 p-ISSN : 2541-044X

Based on the results of research conducted by the author, it was found that the location of the crime of theft was dominated by parking areas with a percentage of 44.9%. This shows that the parking area is a location that is prone to criminal acts of theft, perhaps due to the lack of supervision and the presence of vehicle owners in the area. Furthermore, theft in the house is also a place where criminal acts of theft often occur with a percentage of 22.4%. This shows that houses are the main target for criminals to commit theft. The high percentage of theft in the house may be caused by negligence in securing the house, lack of public awareness to protect their property, or the existence of security holes that are exploited by criminals. In addition, the highway is also a place that is prone to criminal acts of theft with a percentage of 20.4%. This is understandable considering that highways are often places of heavy and busy traffic, thus creating opportunities for criminals to carry out their actions. Meanwhile, the home page is a place where criminal acts of theft are rare with a percentage of 12.2%. Perhaps because of the presence of a home page that tends to be more protected and difficult for strangers to access, making it a relatively safe place from criminal acts of theft.

These findings provide important information in efforts to handle and prevent criminal acts of theft. The focus on security and surveillance in vulnerable areas such as parking areas, houses and roads needs to be increased in order to reduce the number of cases of theft and improve public security.

Victim Profile Analysis

The profile characteristics of the victims in this study are divided into several parts, namely based on the survey results, the sex of the victims of the crime of theft is mostly dominated by women with a percentage of 59.2%. Based on survey results, late teens are the most common age category for theft victims among others with a percentage of 38.8%. Based on the survey results, people who work as sales entrepreneurs in shops and markets are victims of theft crimes with the highest percentage, namely 44.9%. The urban zone is the area with the most cases of theft crimes based on survey results with a total of 42.9%. According to the survey results, the most frequent criminal acts of theft occurred in parking areas with cases reaching 44.9%. The urban zone is the area where the most victims of theft crime live based on the survey results with a total of 49%. Based on the results of the survey, early morning is the time most prone to criminal acts of theft with a percentage of 38.8%. Sundays are days that are prone to criminal acts of theft, based on survey results there were 30.6% of cases occurring on Sundays. Based on the survey results, it was proven that the perpetrators of criminal acts of theft used a lot of tools when carrying out their actions, with a percentage of up to 93.9%. Items that are targeted by the perpetrators of theft are usually movable or immovable goods with a percentage of 53.1%; and followed by moving goods as much as 40.8%. Most of the victims who experienced the crime of theft suffered losses above IDR 2,500,000 with a total of around 93.9%. Most of the perpetrators of the crime of theft carried out their actions by damaging the locks with a total of 81.6%. Based on survey results, the article most frequently violated in the crime of theft is article 363 of the Criminal Code with 93.9%.

CONCLUSION

The conclusions based on the results of research on the prediction of criminal acts of theft using the K-Nearest Neighbor Classifier Machine Learning method by Bandar Lampung Police investigators in the police 4.0 era are as follows:

1. The results of the implementation and evaluation of predictions using the k-NN algorithm with Rapidminer starting from the process of cleansing data, retrieving data, split data, data training (k-NN), data testing (Apply Model), and performance show that the data comparison training and testing respectively 9:1, all K values of K = 1, 3, 5, 7 and 9 have the same accuracy value of 99.40% which is due to the too small amount of testing data, namely 10%. Then in the comparison of training and testing data respectively 8:2, the value of K that has the best level of accuracy is $3 \leq K \leq 5$ with an accuracy rate of 99.10%. Then in the comparison of training and testing data respectively 7:3, the value of K which has the best level of accuracy is K=3 with an accuracy rate of 99.20%. Finally, on a comparison of training and testing data respectively 6:4, the value of K that has the best level of accuracy is K=5 or K=9 with an accuracy rate of 98.80%.

*name of corresponding author

2. The results of the comparison of the prediction results with the k-NN algorithm using Rapidminer show that the comparison of training data and successive testing data which has the best accuracy value is 9:1 with the highest accuracy rate being 99.40%. However, because the amount of testing data is too small, it creates an anomaly in the prediction results where all prediction results with variations in K values have the same level of accuracy. Therefore, a comparison of training data and testing data is taken which has the second highest accuracy level, namely 7:3 with the highest accuracy level of 99.20% with a value of K = 3.
3. The results of the analysis and profiling show that the most common type of theft crime from 2019-2022 was the crime of theft with weighting (CURAT) of 62.20%, followed by 25.60% of cases which were ordinary theft cases (CUBIS). and 12.20% of other cases were cases of theft with violence (CURAS). The majority of victims who experience criminal acts of theft are adult men who live unemployed living in urban areas. The theft occurred in an urban area parking area, on Monday morning. The perpetrator used a tool when carrying out the action, targeting moving objects and causing great losses, the modus operandi used was damaging the lock with the type of theft being weighted theft (CURAT) with article 363 of the Criminal Code.

Recommendations for this research for further research are expected in future research, it can also be discussed more broadly in other articles regarding the crime of theft. Future research is expected to produce more in-depth and accurate predictions by utilizing data that has an even longer time span. The sample data used in future research is expected to be able to predict criminal acts of theft in more detail down to the profile of the perpetrator and also in future research it is expected to be able to use the collaboration of several prediction methods in order to produce much better data predictions and data classification..

REFERENCES

Badrul Mohammad. (2016). Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan. *Jurnal Pilar Nusa Mandiri*, *12*(2), 121–130.

Brodley, C. E., Lane, T., & Stough, T. M. (1999). Knowledge discovery and data mining. *American Scientist*, *87*(1), 54–61. https://doi.org/10.1511/1999.16.807

Chazawi, A. (2002). *Pelajaran Hukum Pidana*. PT Raja Grafindo Persada.

Humas Polri. (2023). https://humas.polri.go.id/tugas-dan-fungsi/

Hakim, L. A. R., Rizal, A. A., & Ratnasari, D. (2019). Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN). *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *1*(1), 30–36. https://doi.org/10.35746/jtim.v1i1.11

Larose D, T. (2005). *Discovering knowledge in data : an introduction to data mining*. Jhon Wiley & Sons Inc.

Rachmadie, D. T. (2020). PENERAPAN ARTIFICIAL INTELLIGENCE PADA TINDAK PIDANA MALWARE DAN PENYIMPANGANNYA BERDASARKAN UNDANG-UNDANG NOMOR 19 TAHUN 2016. *Jurnal Hukum Universitas Surakarta*.

Umair, A., Sarfraz, M. S., Ahmad, M., Habib, U., Ullah, M. H., & Mazzara, M. (2020). Spatiotemporal analysis of web news archives for crime prediction. *Applied Sciences (Switzerland)*, *10*(22), 1–16. https://doi.org/10.3390/app10228220

Umoh, U., Umoh, U., Eyoh, I., & Nyoho, E. (2021). *A Comparative Analysis of k-Nearest Neighbor, Support Vector Machine and Random Forest for Crime Event Type Prediction*.

*name of corresponding author