Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

Prediction of Theft with Machine Learning Technology at Police Station

Aditya Hadmanto*, Jarot Prianggono

Ilmu Kepolisian, Administrasi Kepolisian, Sekolah Tinggi Ilmu Kepolisian, Jakarta, Indonesia
Email: [1,*]rockassmeteor@gmail.com, [2]komputerstik@gmail.com
Correspondence Author Email: rockassmeteor@gmail.com
Submitted: **24/04/2024**; Accepted: **30/06/2024**; Published**: 30/06/2024**

Abstract−This study originated from the increase in theft cases in the jurisdiction of Banjarbaru District Police which resulted in material and psychological losses for victims and disturbed the overall sense of security of the community. The research aims to develop a method that can assist the police in preventing and tackling theft crimes more effectively using *machine learning* algorithms. Research methods include research design, quantitative approach, and data collection and analysis techniques. The data analyzed included various categories of relevant information, such as the victim's gender, age, occupation, location of the incident, as well as details related to the modus operandi and losses suffered by the victim. The main data used is data on victims of theft crimes in the Banjarbaru Police jurisdiction during the 2019-2023 period. Data collection was carried out using primary data available from Min Ops Reskrim Polresta Banjarbaru. using the *K-Nearest Neighbor* (KNN) and *Naïve Bayes* (NB) algorithms to process historical data on theft crimes in Banjarbaru. The results reveal the general characteristics of theft cases, including time patterns, locations, and modus operandi, and compare the effectiveness between KNN and NB algorithms in predicting theft crimes. The conclusions emphasize the potential of machine learning in identifying theft patterns and provide recommendations for further development to support better decision-making and planning of crime prevention strategies.

Keywords: KNN; Naive Bayes; Machine Learning; District Police; Theft

1. INTRODUCTION

Criminal offenses, which are acts prohibited by law and punishable by punishment, display various types based on the perpetrator, motive, impact, and scope [1]. Types of criminal acts include blue collar, white collar, victimless, organized, corporate, digital, transnational, and international crimes, each with its characteristics, modus operandi, and law enforcement challenges [2]. In particular, the crime of theft is an important highlight because it often occurs in the community, causing losses to victims and society in general [3]. Data from the Indonesian National Police recorded that in 2022 there were 1,021,788 theft cases across Indonesia, with an average of 2,799 cases per day. The types of theft vary, including common theft, aggravated theft, petty theft, violent theft, and family theft. All of this becomes the focus of analysis and research in an effort to understand, analyze, and develop criminal law enforcement in accordance with the values of Pancasila and the 1945 Constitution, as reflected in the *Cambridge Crime Harm Index* (CCHI) with the form of a given graph.

The *Cambridge Crime Harm Index* (CCHI) graph above provides an alarming visualization of the theft crime situation in Indonesia, specifically from 2019 to 2023. After recording a sharp decline in CCHI from 2019 to 2020, we see a significant increase in 2021, reflecting the national trend as reported by the Indonesian National Police with 1,021,788 theft cases in 2022. This shows that, despite annual fluctuations, theft-in its various forms, ranging from simple theft to violent theft-remains a persistent problem that has a major impact on society [4]. This phenomenon, which is reflected in the sharp spike in the CCHI graph, combines with local data from the Banjarbaru Resort Police that shows an upward trend in theft cases in the area. This confirms the importance of patrol activities by the police as one of the prevention and control strategies. This strategy, along with a strict law enforcement approach and community-based crime prevention programs, is important to combat the growing trend of theft, lest it develop into a larger and more difficult problem to control.

Some of the factors that can lead to an increase in theft in the Banjarbaru area include economic factors, social factors and psychological factors [5]. This economic factor is one of the main factors that encourage someone to commit theft. Difficult economic conditions, high unemployment, low income, and difficulty in meeting needs make a person feel desperate and look for shortcuts to get money in a fast and easy way, namely by stealing. This social factor is also quite important because it relates to the social environment that influences a person's behavior. A social environment that is not conducive, such as promiscuity, drug abuse, gambling, and violence can affect one's attitude and morals [6]. A person involved in a negative social environment can be easily influenced to commit unlawful acts, including theft. In addition, the lack of attention and supervision from family, community and government can also make a person feel uncontrolled and free to do anything, including theft. In addition to economic and social factors, psychological factors are also a central factor that is quite influential. This factor relates to a person's mental and mental state. A person with a mental disorder, such as depression, stress, trauma or kleptomania may have an irresistible urge to steal. This can be caused by a variety of things, such as life pressures, personal conflicts, or bad experiences in the past. Someone who is mentally ill or mentally ill needs more serious help and treatment from medical experts and psychologists and is not suitable if they have to be convicted [7].

In 2022, theft was one of the most prevalent criminal offenses, with 33 cases of theft and 39 cases of theft with aggravation. The increase in theft crimes in the Banjarbaru area certainly has a negative impact, especially on the victims themselves. Victims of theft can suffer material losses, namely the loss of stolen goods or property. In addition, victims of theft can also suffer immaterial losses, namely experiencing psychological disorders, such as fear, anger,

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

sadness, or trauma. Victims of theft can also experience a decreased sense of security and comfort in their activities, whether at home, at work, or in public places. Not only the victims but even the perpetrators themselves will also experience adverse effects. The impact on the perpetrators of theft can be criminal punishment, namely in the form of imprisonment, fines, or other penalties in accordance with applicable legal provisions. In addition, perpetrators of theft may also experience social punishment, in the form of rejection, ostracism, or negative stigma from their family, community, or surrounding environment. Perpetrators of theft can also experience a decrease in the quality of life, both in economic, social and moral terms. The community itself will also experience disturbances in order and security due to the increase in theft crimes in the Banjarbaru area. This can reduce trust and solidarity between fellow community members. In addition, people can also experience a decrease in welfare and quality of life due to injustice, inequality, and poverty, which is one of the root causes of theft.

In 2021, Sally Nathalia researched with the title "Tinjauan Kriminologis Terhadap Tindak Pidana Pencurian Motor di Kota Singaraja" and obtained results in the city of Singaraja if presented 90% had a negative impact on motorcycle theft and only 10% were not affected by theft [5].

In 2020, Nadya Septiani conducted a study with the title " Analisis Data Mining Pengelompokkan Kasus Tindak Kejahatan Yang Terjadi Di Kecamatan Medan Polonia Dengan Menggunakan Metode K-Means Clustering " and obtained the results of crime in Medan Polonia sub-district divided into 3 clusters, 1 drug cluster, 2 theft clusters, 3 gambling clusters, and the most crime is cluster 2 theft with 38 thefts[11].

In 2019, Usep Tatang Suryadi conducted a study with the title " Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means " and obtained the results of the largest theft rate in the cities of Bogor, Sukabumi, Cianjur, Bandung, Garut, Cirebon, Majalengka and Karawang[15].

The increase in theft cases, it is important to be aware of theft cases that occur at the grassroots of society that are not detected and caught. Due to a lack of legal knowledge, some members of the community are reluctant to report cases of theft. In preventing this, it is also necessary to update the Banjarbaru Police in its patrols so that it is maximized in securing its territory. So the purpose of this research is to carry out classification and prediction of the time of theft in Banjarbaru using the *K-Nearest Neighbor (KNN)* and *Naïve Bayes (NB)* machine learning algorithms, find the best model of classification and prediction of theft crimes using machine learning algorithms, and assist Banjarbaru Police in creating effective patrol patterns based on time using theft association analysis.

2. RESEARCH METHODOLOGY

2.1 Research Stages

The dataset used for classification and prediction of theft crimes with machine learning algorithms at Banjarbaru Police Station is a primary dataset or dataset obtained from data on victims of theft crimes in the Banjarbaru Police jurisdiction in the 2019-2023 range. the columns contained in the dataset are 28 columns and 1247 rows. All columns will not be used in the process of classification and prediction of theft crimes with machine learning algorithms, only relevant columns will be used for the process. Before attribute selection is carried out, data outside the 2019-2023 range is first deleted because the focus of the research is on 2019-2023 data.

Figure 1. Research Stages

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

2.2 Machine Learning

At first, the concept of *machine learning* was introduced by several mathematicians in the 1920s, including Adrien Marie Legendre, Thomas Bayes, and Andrey Markov, who explained the basic principles of machine learning itself [8]. *Machine learning* is a branch of artificial intelligence that aims to make computers capable of performing tasks without human intervention. This concept allows computers to learn from existing data and perform evaluations and actions based on that information. Using the fields of statistics, mathematics, and data mining, computers will continue to learn and optimize their results, even able to recognize new data and perform certain tasks without human assistance [9]. One famous example of the application of machine learning is Deep Blue developed by IBM in 1996, which was able to win chess games against professional players. Today, *machine learning* has been used in various aspects of daily life, such as cell phone unlocking, web browsing, and social media apps. In general, there are two main methods in machine learning, namely supervised learning, which uses supervised learning to identify patterns based on previous learning experiences, and unsupervised learning, which are used to find hidden patterns in unlabeled.

2.3 Data Mining

Data mining is the process of extracting information from a dataset using specialized methods. This process involves a combination of statistics, math, artificial intelligence, and machine learning to extract and identify useful data in large databases [11]. The quality of the information produced can be measured based on accuracy, completeness, consistency, timeliness, trustworthiness, and interpretability [12]. Types of data mining include description, estimation, prediction, classification, clustering, and association, with common functions such as clustering, pattern discovery, prediction, and information mapping [13]. There are several categories of data mining applications, including predictive and descriptive, each of which aims to estimate attribute values and find relevant patterns in the data [14]. Data mining classification techniques, such as *K-Nearest Neighbor (KNN)* and *Naïve Bayes*, are used to recognize new patterns in data that have not been studied before [15]. Data mining evaluation is important to measure the performance of the model that has been created, using metrics such as *accuracy*, *precision*, *recall*, and *f1-score* based on the *confusion matrix* [12]. As such, *data mining* makes significant contributions to understanding data patterns, predicting behavior, and supporting effective decision-making in a variety of contexts, ranging from business to academic research [16].

2.4 Criminology Theory

Criminology is a branch of science that studies the phenomenon of crime, its causes, and the behavior of criminals who violate the norms of society [5]. The aim is to scientifically understand the crime by using data, patterns and related factors, and formulate solutions to overcome it. Criminology provides guidance on how to effectively reduce and prevent crime. Criminology examines the characteristics of crime, the factors that cause crime, the evolution of criminal law, the profile of individual criminals, and the impact on social change. One relevant theory is Differential Association theory, which explains that crime is learned through interaction with other individuals in a close social environment, involving the teaching of crime techniques and their motives and justifications [6].

3. RESULT AND DISCUSSION

Banjarbaru City in South Kalimantan has unique and diverse geographical characteristics. The city lies on a low-lying plain with some undulating areas and hills. The elevation of the region varies, giving different landscapes in different parts of the city. Banjarbaru is also surrounded by green areas and agricultural land, indicating the fertility of the land and the richness of natural resources. Small rivers flow through the city, contributing to drainage and irrigation systems that support agricultural activities. The population as of 2017 is 258,753. While the population density is 848 people/Km^2.

3.1 Description of Characteristics of Victims of Theft

Before analyzing the data, cleaning and normalization are first carried out so that the data can be analyzed more easily. The data resulting from the pouring of the Police Report after the cleaning and normalization process shows the number of theft crimes throughout 2019-2023 amounted to 1242 cases. The total data of theft crimes in Banjarbaru City for each year is shown in Table 1 as follows.

Table 1. Number of cases

No	Years	Total of Cases
1	2019	412
2	2020	122
3	2021	228
4	2022	241
5	2021	239
Total		1242

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

The highest number of theft cases in Banjarbaru City occurred in 2019 with a total of 412 cases and 2019 contributed 33% of cases.

Table 2. Data on victim's gender

No	Gender of Victim	Years					Total
		2019	2020	2021	2022	2023	
1	Males	236	68	138	145	141	**728**
2	Females	176	54	90	96	98	**514**
3	**Total**	**412**	**122**	**228**	**241**	**239**	**1242**

Based on the information listed in Table 2, it can be concluded that victims of theft in Banjarbaru City from 2019 to 2023 were predominantly male, with a total of 728 cases, which accounted for 59% of the total cases. In addition, the data also shows that victims of theft are generally in the late adolescent and late adult age ranges, with a total of 395 and 355 cases respectively. The percentage of victims in the late adolescent age range was 32%, while victims in the late adult age range reached 29%. The occupations of victims of theft in Kota Banjarbaru also show significant variations, ranging from private employees, housewives, to civil servants. However, the majority of victims worked in the private sector or were associated with private companies, with a total of 528 cases, or around 42%. In addition, in terms of the location of the theft crime, there is some data that lists places outside the sub-districts in Kota Banjarbaru.

Table 3. Subdistrict crime scene data

No	Subdistrict Crime Scene	Years					Total
		2019	2020	2021	2022	2023	
1	Banjarbaru Utara	120	48	65	77	63	373
2	Banjarbaru Selatan	151	49	74	76	74	424
3	Landasan Ulin	115	16	63	59	58	311
4	Cempaka	20	5	19	15	17	76
5	Liang Anggang	3	3	5	13	18	42
6	Others	3	1	2	1	9	16
	Total	412	122	228	241	239	1242

According to the data presented in Table 3, Kecamatan Banjarbaru Selatan mencatat total 424 kasus, atau sekitar 34%, it can be seen that the sub-districts in Banjarbaru City that experienced the most theft crimes from 2019 to 2023 were South and North Banjarbaru Sub-districts. While Banjarbaru Utara Sub-district recorded 373 cases, or about 29%. Data on the address where thefts occurred showed significant variation, including roads, schools and residential areas. Therefore, this data needs to be simplified before further analysis. The results of the simplification show that around 522 cases of theft crimes occurred around the highway, which accounted for 42%. In terms of evidence found, electronic goods were the most common items found in theft crimes in Banjarbaru City, with a total of 509 cases, or approximately 41%. The majority of losses suffered by victims each year exceeded 2.5 million rupiah, with a total of 779 cases, or around 63%. The most common modus operandi used in the crime of theft in Banjarbaru City is by taking goods, which recorded a total of 586 cases, or approximately 47%.

Table 4. Data on types of crime

No	Type of punishment	Years					Total
		2019	2020	2021	2022	2023	
1	Curat	213	74	116	106	66	**575**
2	Curbis	197	43	111	134	168	**653**
3	Curas	2	5	1	1	5	**14**
	Total	**412**	**122**	**228**	**241**	**239**	**1242**

From the information shown in Table 4, it can be concluded that the most common type of theft crime in Banjarbaru City from 2019 to 2023 is Curbis, with a total of 653 cases, or around 53%. The time of day when thefts occurred most commonly occurred in the morning, with a total of 350 cases, or around 28%.

3.2 Classification and Prediction of Theft Crimes with Machine learning Algorithms

a. Dataset Preparation

The dataset used for classification and prediction of theft crimes with *machine learning* algorithms at Banjarbaru Police Station is a primary *dataset* or *dataset* obtained from data on victims of theft crimes in the Banjarbaru Police jurisdiction in the 2019-2023 range. Data outside this range will not be used. Then for the columns contained in the *dataset*, there are 28 columns and 1247 rows. All columns will not be used in the process of Classification and prediction of theft crimes with *machine learning* algorithms, only relevant columns will be used

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

for the process. Before the selection of attributes, data outside the 2019-2023 range was first deleted because the focus of the research was on 2019-2023 data.

b. Selection of attributes

Attribute selection is the stage of selecting classes and features in the dataset columns. For classes, we used data on the time of occurrence with a total of 5 classes, namely, early morning, morning, night, afternoon and evening. Then for the features that will be used are features that are relevant to the research topic raised, namely, sub-district, address crime scene, and day.

c. Preprocessing Data

The *preprocessing* stage is carried out by deleting empty data and normalizing data for data that does not match. First, empty data is checked in the columns that have been previously selected as classes and features, namely time, sub-district, address crime scene, and day. If there is no empty data in the data that will be used for classification and prediction. So there is no need to delete empty data. After the blank data is checked, then normalize the data in each column. Before normalization is carried out, the value is changed to *lower case*, then normalization is carried out. The time and day columns do not need to be normalized because the values in both columns are as required. The result of *preprocessing* is clean and structured data generated from prime data processing with 4 columns and 1242 rows. So using the results of *preprocessing* data can be done training or modeling *machine learning.*

	waktu2	kecamatan	hari	tkp_alamat
0	sore	banjarbaru_utara	selasa	jalan_raya
1	pagi	banjarbaru_selatan	jumat	masjid
2	dini_hari	banjarbaru_selatan	minggu	jalan_raya
3	malam	cempaka	senin	jalan_raya
4	siang	landasan_ulin	rabu	perumahan

Figure 2. Preprocessing

d. K-Nearest Neighbor (KNN)

The implementation of the K-Nearest Neighbor or KNN algorithm uses the default k value in the sklearn library which is k = 5. Then for the division of data for training and testing will be carried out into 5 predetermined ratios. Umair explained that the accuracy of prediction with the KNN algorithm will increase at K values that are not even or odd and eventually reach the maximum value at k = 9. Based on Umair's research, the researcher took the value of K = 1,3,5,7,9. Then each ratio scenario will be compared weighted average for *accuracy*, *precision*, *recall*, and *f1-score* values.

1. Training Data Ratio 0.5 and Test Data 0.5

Table 5. Implementation results of each K value with a ratio of 0.5 : 0.5

Nilai K	Accuracy	Precision	Recall	F1-Score
1	**0.87**	**0.87**	**0.87**	**0.87**
3	0.80	0.81	0.80	0.80
5	0.80	0.81	0.80	0.80
7	0.79	0.80	0.79	0.79

From the results of applying each k value with a training and test data division ratio of 0.5:0.5, as shown in Table 5, it can be concluded that the use of k = 1 produces the best model.Dengan rasio tersebut, KNN model achieved accuracy, precision, recall, and f1-score of 0.87. The implementation with k=1 shows good consistency in model performance, with values stabilizing at 0.87 for each evaluation metric. However, when odd k values from 3 to 9 are applied, there is a significant decrease in accuracy, precision, recall, and f1-score. This decrease ranges from 0.07 to 0.08 for accuracy, 0.06 to 0.07 for precision, and 0.07 to 0.08 for recall and f1-score. The use of k=1 makes the KNN model more adaptive to the data because it only considers the closest neighbors to make predictions. This allows the model to capture more specific patterns in the training data. In the case of small datasets, using larger k values such as 3, 5, 7, or 9 may make the model too complex and prone to overfitting. With k=1, the model tends to be simpler and has a lower risk of overfitting. Therefore, the implementation results with the value of k = 1 show that the KNN model is better in the training data and test data division ratio of 0.5: 0.5.

2. Training Data Ratio 0.6 and Test Data 0.4

Table 6. Implementation results of each K value with a ratio of 0.6 : 0.4

Nilai K	Accuracy	Precision	Recall	F1-Score
1	**0.89**	**0.89**	**0.89**	**0.89**
3	0.84	0.86	0.84	0.84
5	0.84	0.85	0.84	0.84

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

7	0.87	0.87	0.87	0.87
9	0.83	0.84	0.83	0.83

From the results of applying each k value with a training data and test data division ratio of 0.6: 0.4, as shown in Table 6, the conclusion is that the use of k=1 produces the best model. With these ratios, the KNN model achieved accuracy, precision, recall, and f1-score of 0.89, showing good consistency in model performance. However, when odd k values from 3 to 5 are applied, there is a decrease in accuracy, precision, recall, and f1-score. After that, the k=7 value resulted in an improvement, with accuracy, precision, recall, and f1-score stabilizing at 0.87. However, the values decrease again for k=9. However, this decrease is still smaller than what happens when the value of k=1. Just like in the 0.5:0.5 ratio, in the case of datasets that tend to be small, using larger k values such as 3, 5, 7, or 9 can make the model too complex and prone to overfitting. With k=1, the model tends to be simpler and has a lower risk of overfitting. By considering these factors, the implementation with k=1 shows that the KNN model is better in the training and test data division ratio of 0.6:0.4.

3. Training Data Ratio 0.7 and Test Data 0.3

Table 7. Implementation results of each K value with a ratio of 0.7 : 0.3

Nilai K	Accuracy	Precision	Recall	F1-Score
1	**0.91**	**0.91**	**0.91**	**0.91**
3	0.86	0.88	0.86	0.86
5	0.82	0.84	0.82	0.82
7	0.82	0.84	0.82	0.82
9	0.80	0.81	0.80	0.80

From the implementation results with each k value and training and test data division ratio of 0.7:0.3, as noted in Table 7, the conclusion is that a value of k=1 produces the best and most stable model. Using this k value, the KNN model achieved accuracy, precision, recall, and f1-score of 0.91, showing good consistency in model performance. However, when odd k values from 3 to 5 are applied, there is a gradual decrease in accuracy, precision, recall, and f1-score. The value of k=7 has the same performance as k=5, but again there is a decrease for k=9. However, the decrease in value is still greater than when the value of k = 1. As with the 0.5:0.5 and 0.6:0.4 ratios, in the case of datasets that tend to be small, using larger k values such as 3, 5, 7, or 9 can make the model too complex and prone to overfitting. By using a value of k=1, the KNN model tends to be simpler and has a lower risk of overfitting. Therefore, the implementation result with k=1 shows that the KNN model is better in the training and test data division ratio of 0.7:0.3 than using the highest k value, which is k=9.

4. Ratio of Training Data 0.8 and Test Data 0.2

Table 8. Implementation results of each K value with a ratio of 0.8: 0.2

Nilai K	Accuracy	Precision	Recall	F1-Score
1	**0.91**	**0.91**	**0.91**	**0.91**
3	0.88	0.89	0.88	0.88
5	0.87	0.88	0.87	0.87
7	0.86	0.86	0.86	0.86
9	0.89	0.89	0.89	0.89

From the implementation results of various k values with a ratio of 0.8:0.2, as shown in Table 8, it can be concluded that the value of k = 1 produces the best and stable model. By dividing the training data by 0.8 (993 data) and the test data by 0.2 (249 data), the KNN model achieved accuracy, precision, recall, and f1-score of 0.91, showing good consistency in model performance. However, when odd k values from 3 to 7 are applied, there is a gradual decrease in accuracy, precision, recall, and f1-score. However, there is a significant increase when the value of k=9. Nonetheless, the decrease remains relatively small, ranging from 0.02 to 0.05, for each evaluation metric. As in the previous ratio, in the case of datasets that tend to be small, using larger k values such as 3, 5, 7, or 9 can make the model too complex. Therefore, the implementation result with k = 1 shows an advantage in the training and test data division ratio of 0.8:0.2. In addition, the difference between the highest and lowest values for each evaluation metric narrows and stabilizes around 0.05.

5. Training Data Ratio 0.9 and Test Data 0.1

Table 9. Implementation results of each K value with a ratio of 0.9 : 0.1

Nilai K	Accuracy	Precision	Recall	F1-Score
1	0.91	0.91	0.91	0.91
3	**0.93**	**0.93**	**0.93**	**0.93**
5	0.91	0.91	0.91	0.91
7	0.91	0.91	0.91	0.91

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

From the results of applying various k values with a ratio of 0.9:0.1, as shown in Table 9, The conclusion is that the value of k = 1 produces the best and stable model. By dividing the training data by 0.9 (1117 data) and the test data by 0.1 (125 data), the KNN model achieved accuracy, precision, recall, and f1-score of 0.93, showing good consistency in model performance. The use of the 0.9:0.1 ratio makes the implementation of other k values also show stability in performance. This is due to the larger amount and variety of training data, which causes the model to perform better. However, it should be noted that a value of k=1 shows lower performance compared to the previous ratio. This is due to the increasing complexity as the amount of training data increases. Using k values that are too complex, such as k=5 to k=9, results in a decrease in performance. However, the decrease is still quite small, ranging from 0.02 to 0.04 for each evaluation metric. Thus, the implementation results with the k=3 value show that the KNN model is better in the training and test data division ratio of 0.9:0.1, because the model is not too simple but also not too complex.

e. Naive Bayes

The specific Naïve Bayes algorithm implementation uses the Gaussian Naïve Bayes method. GaussianNB is often chosen because of the assumption that features follow a normal distribution, which is quite common in many cases. Then the division of data for training and testing will be carried out into 5 predetermined ratios. Then each ratio scenario will be compared to the weighted average for accuracy, precision, recall, and f1-score values.

Table 10. Implementation results of each K value with a ratio of 0.9 : 0.1

Rasio	Accuracy	Precision	Recall	F1-Score
0.5:0.5	**0.31**	**0.35**	**0.31**	**0.30**
0.6:0.4	0.28	0.31	0.28	0.26
0.7:0.3	0.27	0.25	0.27	0.24
0.8:0.2	0.29	0.29	0.29	0.27
0.9:0.1	0.30	0.31	0.30	0.29

From the results of the implementation of the Naïve Bayes model with 5 ratios of division of training data and test data shown in Table 10, it can be concluded that the implementation of the division ratio of training data 0.5 (621 data) and test data 0.5 (621 data) produces the best and stable model with accuracy 0.31, precision 0.35 recall 0.31 and f1-score 0.30.

3.3 Evaluation

Based on the graphic shown in Figure 3, The implementation of 0.5:0.5 ratio resulted in the highest accuracy, precision, recall, and f1-score, namely accuracy 0.31, precision 0.35 recall 0.31 and f1-score 0.30. However, overall the Naïve Bayes model is more specific to the Gaussian Naïve Bayes method, less able to classify crimes using the given features because the results of accuracy, precision, recall, and f1-score can be said to be poor overall for each ratio of training data and test data. The results of classification using this method are fairly far when compared to the results of the KNN method. This may occur due to the different characteristics between the two methods. The NB method applies the assumption that each feature is independent of each other, although in the real world, features can be interrelated. This can be a significant source of errors, especially if the features have strong dependencies.

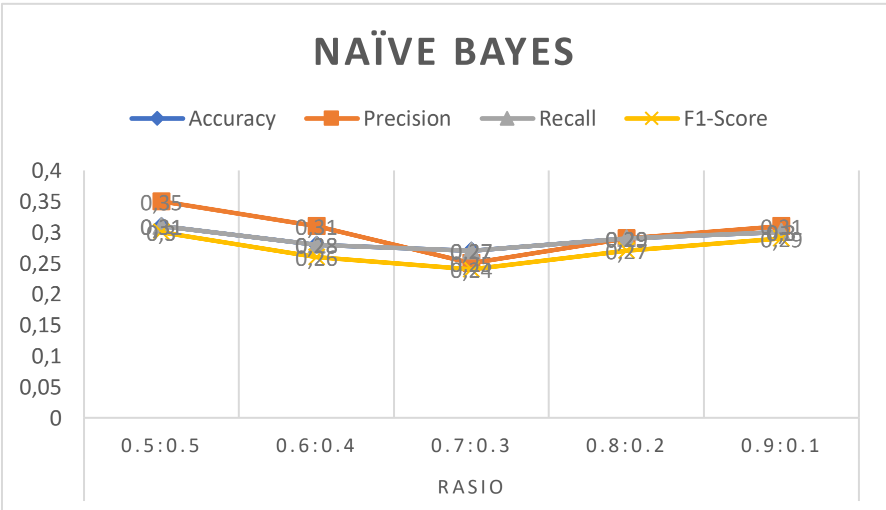

Figure 3. Evaluation NB ratio

a. *Machine Learning Model Evaluation*

Machine learning model evaluation is done by comparing the weighted average results for the accuracy value, precision, recall, and f1-score from each experiment that has been done on the K-Nearest Neighbor algorithm (the

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

best model from each implementation of the k value in each ratio) and Naïve Baye. A summary of the results of the machine learning algorithm testing that has been carried out is shown in Table 11.

Table 11. Summary of machine learning algorithm test results

Algoritma	Rasio	*Accuracy*	*precision*	*recall*	*f1-score*
KNN	0.5:0.5 k=1	0.87	0.87	0.87	0.87
	0.6:0.4 k=1	0.89	0.89	0.89	0.89
	0.7:0.3 k=1	0.91	0.91	0.91	0.91
	0.8:0.2 k=1	0.91	0.91	0.91	0.91
	0.9:0.1 k=3	**0.93**	**0.93**	**0.93**	**0.93**
NB	0.5:0.5	0.31	0.35	0.31	0.30
	0.6:0.4	0.28	0.31	0.28	0.26
	0.7:0.3	0.27	0.25	0.27	0.24
	0.8:0.2	0.29	0.29	0.29	0.27
	0.9:0.1	0.30	0.31	0.30	0.29

Based on the test results shown in Table 11, it can be concluded that the best model that can be implemented to predict theft crimes in Banjarbaru City using 3 features is the sub-district crime scene, the crime scene of the address and day of the event and the class using the time of the event is the KNN algorithm using a ratio of 0.9 (1117 data) as training data and 0.1 (125 data) as test data by implementing a value of k = 3. The best method scored accuracy 0.93, precision 0.93, recall 0.93, and f1-score 0.93. Using the best model, predictions can be made using inputs of sub-district crime scene, address crime scene and day of the incident. An example of the best model prediction using inputs of Landasan Ulin sub-district crime scene, residential address crime scene, and Wednesday is shown in Figure 4.

```
Hasil prediksi untuk input : perumahan, landasan_ulin, sabtu
Prediksi waktu : [sore] : 100.00 %
Prediksi waktu : [dini_hari] : 0.00 %
Prediksi waktu : [malam] : 0.00 %
Prediksi waktu : [pagi] : 0.00 %
Prediksi waktu : [siang] : 0.00 %
Prediksi waktu terbaik: [sore] dengan confidence 100.00%
```

Figure 4. prediction with confidence

Based on figure 5, for the specified input predicted the best model as afternoon time with a confidence value of 100%. Using the best model, a schedule can be created using the input of the sub-district crime scene, address crime scene and the day of the incident to be predicted.

	kecamatan	hari	lokasi	waktu
49	banjarbaru_selatan	senin	perumahan	siang
91	banjarbaru_selatan	senin	toko, kost, rumah_makan, perkantoran, masjid, ...	sore
7	banjarbaru_selatan	senin	jalan_raya, caffe	malam
50	banjarbaru_selatan	selasa	perumahan	pagi
8	banjarbaru_selatan	selasa	jalan_raya, toko, kost, rumah_makan, perkantor...	malam

Figure 5. Patrolling schedule

4. CONCLUSION

The results of the analysis and discussion of this research provide an in-depth understanding of the pattern of theft crimes in the Banjarbaru District Police jurisdiction. In crime classification and prediction, K-Nearest Neighbor (KNN) and Naïve Bayes (NB) algorithms play an important role by utilizing features such as the location of the incident and the day of the incident. The best model, which is the result of dividing the data with a ratio of 0.9:0.1 and implementing a k value of 3 in the KNN algorithm, is able to provide high accuracy, precision, recall, and f1-score values. Association analysis also provides valuable insights into crime patterns, enabling Banjarbaru Police to design more effective patrol strategies and smarter decision-making in law enforcement. Suggestions for further development include the integration of the analysis results into a decision-making system, the addition of additional features in the dataset, as well as increased cooperation with local communities for more inclusive crime prevention. As such, this research not only provides deep insights into theft crimes, but also provides a foundation for the development of security strategies that are more adaptive and responsive to local crime dynamics.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 550-558
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5107

REFERENCES

[1] T. Hartono, M. A. Lubis, and S. A. Siregar, “Penegakan Hukum Terhadap Tindak Pidana Pencurian Dengan Kekerasan (Studi Pada Kepolisian Resor Kota Besar Medan),” Jurnal Retentum, vol. 3, no. 1, pp. 32–42, 2021, doi: 10.46930/retentum.v3i1.900.

[2] M. T. I. Wahyudin, S. Shafira, F. Putri, and R. S. Putra, “Pengegakan Hukum terhadap Tindak Pidana Pencurian dengan Kekerasan,” Jurnal Edukasi Nonformal, vol. 4, no. 1, pp. 228–238, 2023.

[3] I. A. Sakti, M. Ilyas, and M. Z. Muhdar, “Tinjauan Yuridis Terhadap Tindak Pidana Pencurian dengan Kekerasan,” Qawanin Jurnal Ilmu Hukum, vol. 2, no. 2, pp. 1–12, 2021, doi: 10.58258/jisip.v4i4.1479.

[4] J. C. Putra and D. A. D. Tawang, “TINDAK PIDANA PENCURIAN YANG DILAKUKAN SECARA BERSEKUTU STUDI PUTUSAN NOMOR 203/Pid.B/2021/PN.LLG,” Reformasi Hukum Trisakti, vol. 5, no. 4, pp. 1300–1309, 2023, doi: 10.25105/refor.v5i4.18581.

[5] S. N. Christie, N. P. R. Yuliartini, and D. G. S. Mangku, “Tinjauan Kriminologis Terhadap Tindak Pidana Pencurian Kendaraan Bermotor Di Kota Singaraja,” Journal Komunitas Yustisia Universitas Pendidikan Ganesha Program Studi Ilmu Hukum, vol. 4, no. 1, pp. 119–125, 2021.

[6] A. W. Andani, M. R. Bima, and Sutiawati, “Tinjauan Kriminologi Terhadap Tindak Pidana Pencurian Ternak,” Qawanin Jurnal Ilmu Hukum, vol. 1, no. 1, pp. 1–40, 2020.

[7] Nathalia, G. Angel, and H. S. Muaja, “Tinjauan Kriminologis Terhadap Residivis Tindak Pidana Pencurian,” Lex Privatum, vol. 10, no. 5, pp. 1–8, 2022.

[8] F. A. Ardandy, I. M. Pohan, A. Mitsal, F. Nusantara, and M. D. Ruliansyah, “Perbandingan Algoritma Naive Bayes dan Linear Discriminant Analysis dengan Dataset Car Evaluation,” Jurnal Rekayasa Elektro Sriwijaya, vol. 3, no. 1, pp. 213–217, 2021, doi: 10.36706/jres.v3i1.45.

[9] R. Debnath, R. Bardhan, D. M. Reiner, and J. R. Miller, “Political, economic, social, technological, legal and environmental dimensions of electric vehicle adoption in the United States: A social-media interaction analysis,” Renewable and Sustainable Energy Reviews, vol. 152, p. 111707, 2021, doi: https://doi.org/10.1016/j.rser.2021.111707.

[10] A. I. Putra and R. R. Santika, “Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering,” Edumatic : Jurnal Pendidikan Informatika, vol. 4, no. 1, pp. 121–130, 2020, doi: 10.29408/edumatic.v4i1.2162.

[11] N. Septiani, K. Erwansyah, and M. G. Suryanata, “Analisis Data Mining Pengelompokkan Kasus Tindak Kejahatan Yang Terjadi Di Kecamatan Medan Polonia Dengan Menggunakan Metode K-Means Clustering,” Jurnal Cyber Tech, vol. 3, no. 2, pp. 1–13, 2020.

[12] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, “Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa,” IJCIT (Indonesian Journal on Computer and Information Technology), vol. 6, no. 2, pp. 118–127, 2021.

[13] N. C. Siregar, R. R. A. Siregar, and M. Y. D. Sudirman, “Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ),” Jurnal Teknologia Aliansi Perguruan Tinggi (APERTI) BUMN, vol. 3, no. 1, pp. 102–110, 2020.

[14] Y. S. T. Allo, V. Sofica, N. Hasan, and M. Septiani, “Penggunaan Metode Naïve Bayes Dalam Mengklasifikasi Pengangguran Pada Desa Bojong Kulur,” Bianglala Informatika, vol. 10, no. 1, pp. 30–35, 2022, doi: 10.31294/bi.v10i1.12333.

[15] U. T. Suryadi and Y. Supriatna, “Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means,” Jurnal Teknologi dan Komunikasi STMIK Subang, vol. 12, no. 1, pp. 15–27, 2019, doi: 10.47561/a.v12i1.147.

[16] D. Ulhaq, N. Suarna, and G. Dwilestari, “Klasifikasi Berita Kriminal Menggunakan Algoritma Naïve Bayes Berbasis PSO,” Informatics for Educators and Professionals, vol. 6, no. 2, pp. 12–21, 2022.

[17] V Jackins, S Vimal, “Al-based smart prediction of clinical disease using random forest classifier and Naive Bayes”, 2021.

[18] PR Putri Ramayanti “Perbandingan Algoritma Naive Bayes dan Svm Untuk analisis Penyalahgunaan Kejahatan Carding”, 2023.

[19] IC KAMAGI, AM Adrian and Apriandy“Klasifikasi kejahatan trafficking Pada Sosial Media Facebook Menggunakan Algoritma Naive Bayes”, 2021.

[20] A Saputra, FN Hasan “Analisis Sentimen Terhadap Aplikasi Coffee Meets Bagel Dengan Algoritma Naive Bayes Classifier”, 2023