DINASTI RESEARCH
JLPH: Journal of Law, Politic and Humanities
E-ISSN: 2962-2816
P-ISSN: 2747-1985
https://dinastires.org/JLPH dinasti.info@gmail.com +62 811 7404 455

DOI: https://doi.org/10.38035/jlph.v5i5

Prediction of Kamtibmas Trends in the Jurisdiction of Bireuen Police Resort Using Naïve Bayes and Random Forest

Pringgo Putra[1], Novi Indah Earlyanti[2], Vita Mayastinasari[3], Saut Panggabean Sinaga[4].
[1]Polres Bireuen Polda Aceh, primapringgo18@gmail.com.
[2]Sekolah Tinggi Ilmu Kepolisian, noviindahearlyanti@stik-ptik.ac.id.
[3]Sekolah Tinggi Ilmu Kepolisian, mamayovi@gmail.com.
[4]Sekolah Tinggi Ilmu Kepolisian, beatrichsweet@gmail.com.

Corresponding Author: primapringgo18@gmail.com[1]

Abstract: Predicting trends in Public Security and Order (Kamtibmas) is crucial for supporting strategic decision-making by law enforcement agencies, particularly in regions with dynamic social and political environments such as Bireuen Regency. One of the key challenges is the absence of a data-driven predictive system capable of accurately identifying patterns in Kamtibmas incidents. This study aims to develop a predictive model for Kamtibmas trends within the jurisdiction of the Bireuen Police using Naïve Bayes and Random Forest machine learning algorithms. A quantitative approach is employed, following the Knowledge Discovery in Databases (KDD) methodology, which encompasses data selection, preprocessing, transformation, data mining, evaluation, and interpretation. The dataset, sourced from the daily reports of the Bireuen Police Intelligence Unit from 2021 to 2024, was encoded and normalized across variables such as time, day, month, sub-district, incident category, and reporting unit. The results indicate that the Random Forest algorithm significantly outperforms Naïve Bayes. Using a 90:10 split for training and testing data, Random Forest achieved an accuracy, precision, recall, and F1-score of 98%. In contrast, Naïve Bayes demonstrated lower performance, with accuracy ranging between 42% and 44%. These findings suggest that Random Forest is more effective in capturing complex patterns within Kamtibmas data and has strong potential for implementation as a strategic tool to support crime princidention and public order maintenance efforts in Bireuen Regency.

Keywords: Kamtibmas, Naïve Bayes, Random Forest, Data Mining, Prediction

INTRODUCTION

The implementation of elections serves as an indicator of democratic progress toward a civil society. However, it is often accompanied by social unrest that has the potential to disrupt Public Security and Order (Kamtibmas), particularly in the lead-up to the 2024 simultaneous elections. In this context, the role of the Indonesian National Police (Polri) becomes increasingly critical in maintaining public order, princidenting conflicts, and upholding the rule of law (Purwatiningsih & Polri, 2023:63). As stipulated in the Chief of Police Regulation No. 2 of 2021, the National Police operates through a hierarchical organizational structure extending from the national headquarters to local police stations, with district and municipal-level police units playing a key role in executing policing functions. Among these functions is intelligence gathering, which is essential for detecting and issuing early warnings regarding potential Kamtibmas issues.

The importance of this intelligence function is further emphasized in National Police Decree No. 1 of 2023, which outlines the role of police intelligence in supporting internal security (Kamdagri) by ensuring the stability of Kamtibmas. According to Law No. 2 of 2002, Kamtibmas is defined as a dynamic condition that serves as a cornerstone of successful national development, characterized by the maintenance of public order, legal certainty, and the community's ability to resist security threats (Tampubolon, 2023). Therefore, maintaining the stability of Kamtibmas is a shared responsibility between the community and the government, aimed at fostering a safe and peaceful environment (Rosmaya et al., 2022:1314).

Understanding trends in Public Security and Order (Kamtibmas) serves as a critical foundation for the intelligence function in carrying out early detection and princidention of potential disturbances arising from political, economic, socio-cultural, and security sectors (Poleksosbudkam), as outlined in National Police Regulation No. 1 of 2023. Each of these dimensions plays a strategic role in shaping regional stability, particularly in political contexts such as local or national elections, which are often vulnerable to conflict due to intense competition, dissatisfaction with election outcomes, smear campaigns, and money politics (Purwatiningsih & Polri, 2023:53). In regions with a history of conflict, such as Aceh, political dynamics tend to be more sensitive. This includes Bireuen Regency, which frequently experiences heightened tensions stemming from the activities of former Free Aceh Movement (GAM) sympathizers. These tensions often extend beyond the campaign period, persisting through vote counting and the inauguration process, thereby posing a risk to regional stability (Akbar, 2024). This situation is further complicated by the change in regional leadership during the 2021–2024 period. Following the end of the term of Regent Dr. H. Muzakkar A. Gani in 2021, the regional elections that were originally scheduled for 2022 were postponed to 2024 in accordance with the national policy of simultaneous elections. As a result, Bireuen Regency has been led by an Acting Regent (Penjabat or PJ) appointed by the Minister of Home Affairs. This transitional governance has indirectly influenced policy direction and contributed to shifts in Kamtibmas trends within the region.

In 2022, Aulia Sofyan, Ph.D., who previously served as the Head of the Aceh Provincial Investment and One-Stop Integrated Services Office (DPMPTSP), was officially appointed as the Acting Regent (PJ) of Bireuen on August 15, 2022. His term was extended in August 2023 until 2024, before incidentually being succeeded by Jalaluddin, S.H., M.M., on August 11, 2024. Jalaluddin will serve until the official inauguration of the elected Regent and Deputy Regent following the 2024 Regional Elections. Each leadership transition introduces new policies and approaches to managing Kamtibmas, which inevitably affects regional stability—especially when compounded by external factors such as the Covid-19 pandemic, which has also significantly impacted social conditions and community security. The researcher also conducted observations of the Security Intelligence Unit (Sat Intelkam) of the Bireuen Police to identify obstacles in predicting Kamtibmas trends. The findings revealed several issues, including a poorly organized archiving and intelligence administration system, the common use of copy-paste methods in intelligence product development, unsystematic file storage leading to the risk of data loss, and weaknesses in the intelligence analysis section, which lacks depth and fails to comprehensively incorporate historical context. In addition, interviews with several officers indicated ongoing difficulties in comprehensively mapping Kamtibmas trends across all sub-districts within the Bireuen Police jurisdiction.

Various problems within the reporting and information management system of the Bireuen District Police Intelligence Unit, such as weak administration, irregular data storage, and low-quality intelligence analysis, indicate significant barriers to detecting and predicting potential Kamtibmas vulnerabilities in the region. If left unaddressed, these weaknesses could affect the performance of the Intelligence Unit (Sat Intelkam) in maintaining security stability. As a result, technological approaches such as data mining have become highly relevant and urgent to implement. Data mining is capable of efficiently processing and analyzing large volumes of data to identify strategic patterns, such as mapping the intensity of community activities, identifying incident-prone areas, and pinpointing locations frequently experiencing Kamtibmas disturbances. This approach supports more targeted, data-driven decision-making in princidentive measures. According to Fakhri et al. (2023), data mining is an analytical process that uses statistical methods, artificial intelligence, and machine learning to extract valuable hidden insights from large and complex datasets. In the context of Kamtibmas, analyzing historical data—such as incident types, locations, times, and incident chronologies—greatly aids in mapping evolving security trends. Given that Bireuen Regency has experienced over 1,000 incidents in a single year, it holds great potential for the application of this technique. By conducting in-depth data analysis, law enforcement agencies can adopt more accurate and proactive strategies, thereby enhancing the effectiveness of Kamtibmas management and significantly reducing the risk of security disturbances in the region.

Identifying patterns and trends through data analysis requires selecting the appropriate prediction methods to improve result accuracy. Two algorithms that are particularly relevant in this context are Naïve Bayes and Random Forest. Naïve Bayes is a probabilistic classification method based on Bayes' Theorem, assuming the independence of attributes (Raja Rizki Alanta Nasution & Relita Buaton, 2024), while Random Forest is an ensemble method that combines multiple decision trees to create a more accurate model (Buani, 2024). The reliability of these two algorithms has been demonstrated in various studies over the past five years. For instance, Wibowo & Oesman (2020) found that Naïve Bayes achieved the highest accuracy (65.59%) in predicting crime in Sleman, while Zaidi et al. (2020) recorded a remarkable 99.9% accuracy for Random Forest in crime classification. May et al. (2021) showed an improvement in Naïve Bayes accuracy to 96.6% after applying the Recursive Feature Elimination (RFE) method, closely competing with Random Forest at 96.13%. Other studies, such as those by Febrera & Prianggono (2024), demonstrated 100% accuracy in classifying traffic violations using Naïve Bayes at the Padang Police Station, while Hadmanto & Prianggono (2024) and Widya Bhakti Dira (2024) highlighted the superiority of KNN and ANN in specific cases. Based on these findings, this study employs the Naïve Bayes and Random Forest algorithms to maximize the prediction of Kamtibmas trends within the jurisdiction of the Bireuen District Police. The aim is to provide data-driven recommendations for princidention strategies and enhance law enforcement effectiveness through modern technological approaches, as outlined in the thesis titled: "Prediction of Kamtibmas Trends in the Jurisdiction of the Bireuen District Police Using Naïve Bayes and Random Forest."

METHOD

This study employs a quantitative approach using the Knowledge Discovery in Databases (KDD) method to analyze historical data from the Bireuen District Police intelligence reports in order to predict Kamtibmas trends using two machine learning algorithms: Naïve Bayes and Random Forest. The KDD approach is applied through the stages of data selection, preprocessing, transformation, data mining, pattern evaluation, and result interpretation. The independent variables in this study include the time of the incident, year, month, day, date, sub-district, field, subject, and reporting team, while the dependent variables represent the classification of Kamtibmas domains, such as politics, economy, socio-culture, and security (Lianasari & Ahmadi, 2022). All daily report data from the 2021–2024 period were analyzed using a saturated sampling technique to ensure complete and representative coverage of Kamtibmas conditions in the region (Sugiyono, 2021).

The data collection technique was carried out through the review of Daily Report archival documents and unstructured interviews when necessary. The collected data were validated through verification, screening, procedural checks, completeness assessments, and data processing. The analysis was conducted using Google Colaboratory with Python code, applying the Naïve Bayes and Random Forest models, and evaluating the models using metrics such as accuracy, precision, recall, and F1-score. Naïve Bayes is used to calculate probabilities based on the distribution of data, while Random Forest constructs multiple decision trees to improve classification accuracy for more complex patterns. Through this process, it is hoped that the analysis results will provide data-driven strategic insights for formulating more effective Kamtibmas management policies at the Bireuen District Police.

RESULTS AND DISCUSSION

RESULT

Bireuen Regency, established under Law Number 48 of 1999 and later modified by Law Number 8 of 2000, was created through the expansion of North Aceh Regency. Geographically, the district is situated on the east coast of Aceh Province, covering an area of 1,796.31 km². It is directly bordered by the Straits of Malacca, North Aceh, Central Aceh, Bener Meriah, Pidie, and Pidie Jaya. The district consists of 17 sub-districts and lies on a strategic route across Sumatra, positioning it as a potential center of economic growth in the eastern part of Aceh. According to 2022 data, Bireuen's population reached 443,874, with an average density of 247 people per km². The highest population concentration is in Kota Juang District, while Pandrah has the lowest density. The region's characteristics also influence the dynamics of Kamtibmas, which vary and are reflected in the fluctuations of incidents across sub-districts during the 2021–2024 period.

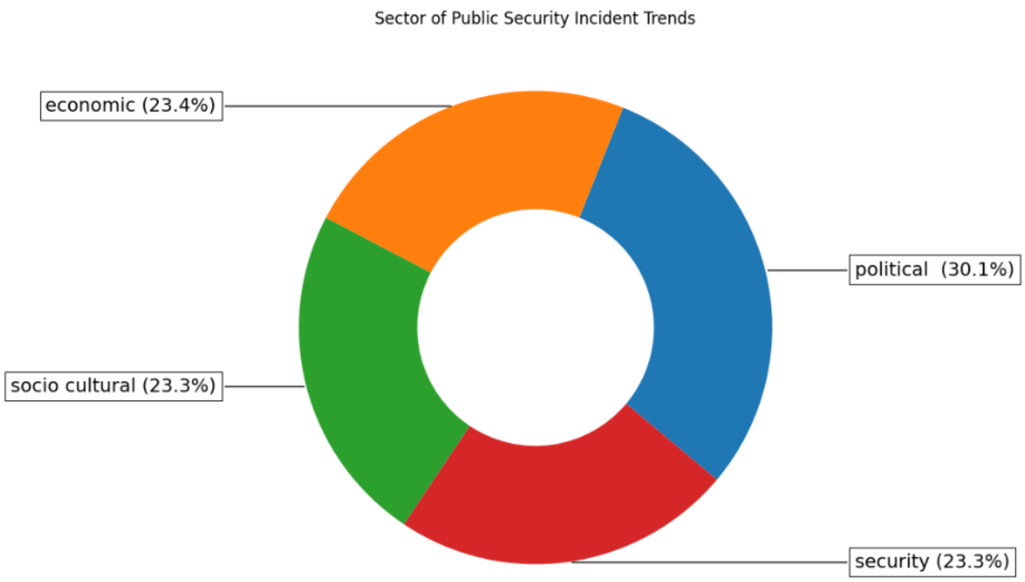


Figure 1. Sector of Public Security Incidents Trends

The Kamtibmas incidents within the jurisdiction of the Bireuen District Police during the 2021–2024 period reveal that the social dynamics in the community can be categorized into four main sectors: politics, economy, socio-culture, and security. Based on data analysis, the political sector emerged as the most dominant, with a total of 1,431 incidents, or approximately 30.1% of the total cases, indicating the significant influence of political issues on the stability of Kamtibmas. This was followed by the economic sector with 1,111 incidents (23.4%), the socio-cultural sector with 1,108 incidents (23.2%), and the security sector with 1,107 incidents (23.2%), all of which made substantial contributions to regional dynamics. This comparison highlights that, although the political sector is the most prominent, incidents in the economic, socio-cultural, and security sectors also play important roles in shaping the overall Kamtibmas trend.

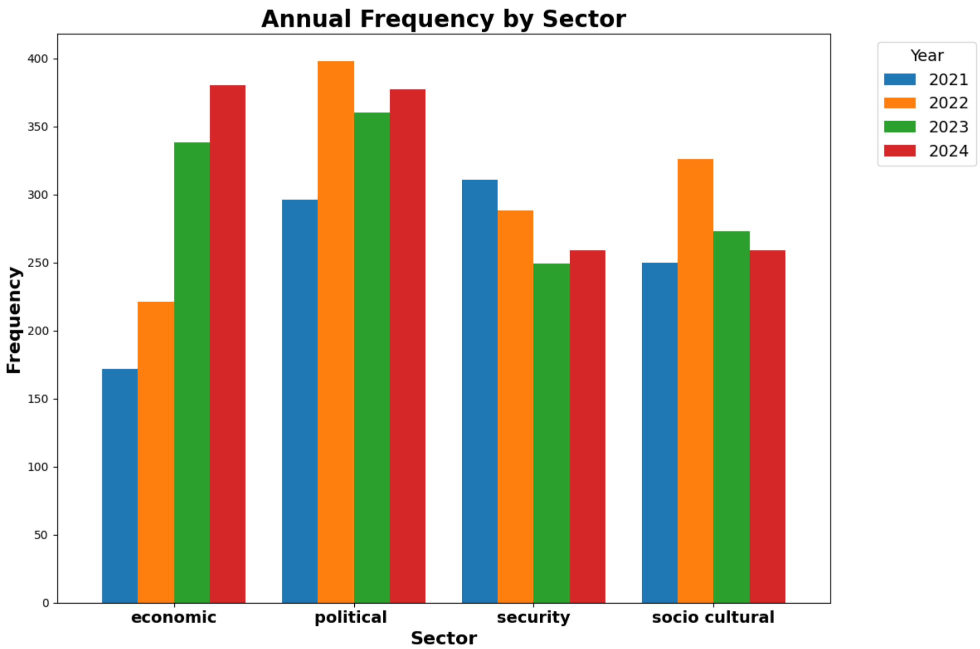


Figure 2. Annual Frequency by Sector

Based on the graph, it can be concluded that the trend of Kamtibmas incidents within the jurisdiction of the Bireuen District Police exhibited varying dynamics across sectors from 2021 to 2024. The political sector recorded the highest frequency in 2022, although it declined in the following year. In contrast, the economic sector showed significant improvements, particularly in 2023 and 2024. Meanwhile, the security and socio-cultural sectors experienced a downward trend after 2022. This pattern suggests that political and economic issues have increasingly taken center stage in the context of Kamtibmas, while security and socio-cultural issues are starting to decrease in frequency. This may indicate a shift in focus or a change in the nature of Kamtibmas disturbances in the community.

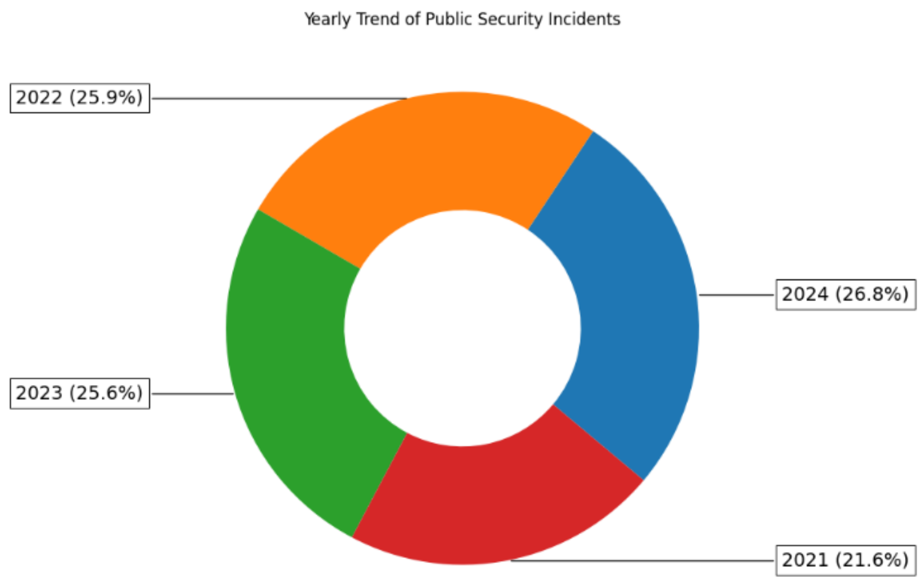


Figure 3. Yearly Trend of Public Security Incidents

Based on the data analyzed from the graphs, the incidence of Kamtibmas within the jurisdiction of the Bireuen District Police shows an increasing trend over the years, with the peak occurring in 2024, which recorded the highest proportion at 26.8%. In 2022 and 2023, the proportions were almost equal, at 25.9% and 25.6%, respectively, while 2021 recorded the lowest frequency at 21.6%. This pattern indicates an escalation in the number of Kamtibmas incidents, particularly in the final year of the analysis period, which may signal the need for increased attention and anticipation of the evolving social dynamics.

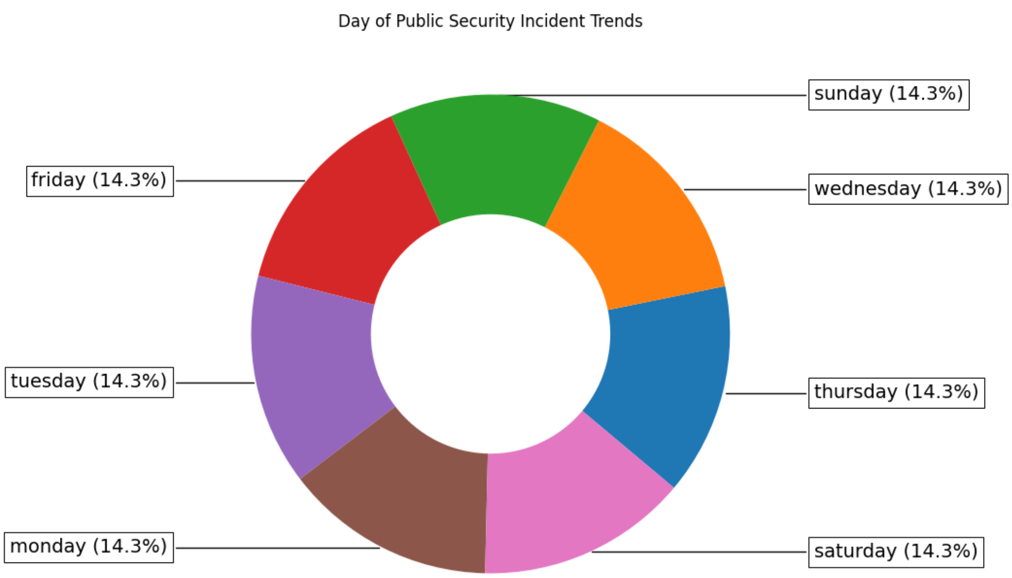


Figure 4. Day of Public Security Incident Trends

Based on the data in the graph in Figure 4, it can be observed that Kamtibmas incidents during the analyzed period are distributed across various days of the week. Additionally, the graph indicates that the day with the highest number of incidents shows an equal proportion, with each day recording 14.3%.

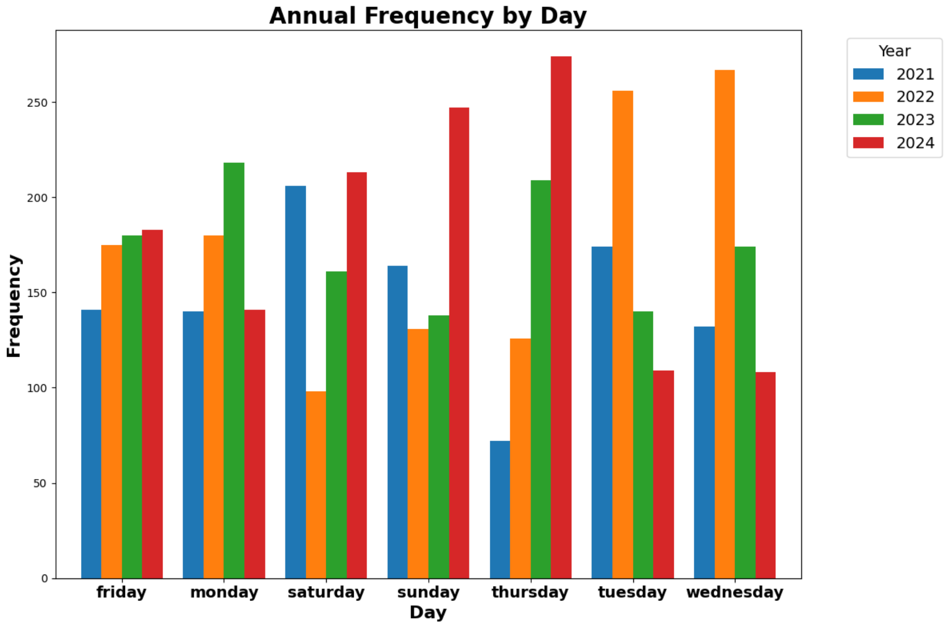


Figure 5. Annual Frequency by Day

Based on the analyzed graph, the frequency of Kamtibmas incidents within the jurisdiction of the Bireuen District Police shows significant variations across different days of the week during the 2021–2024 period. Saturday recorded the highest frequency in 2022, while Wednesday experienced a spike in 2024. Friday and Thursday displayed consistent levels year on year, while Sunday showed a sharp upward trend starting in 2023. In contrast, Mondays and Tuesdays experienced notable fluctuations, with a drastic decline observed in 2024. This pattern reflects the influence of social dynamics that affect the intensity of Kamtibmas incidents on specific days.

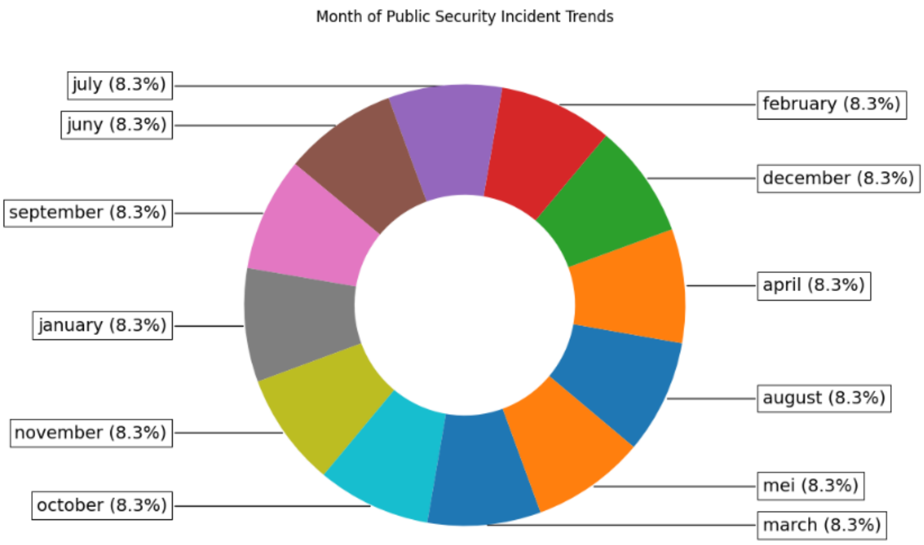


Figure 6. Month of Public Security Incident Trends

Based on the data in the graph on Figure 6, it can be observed that Kamtibmas incidents during the analyzed period are distributed across various months throughout the year. Furthermore, the graph indicates that the month with the highest number of incidents shows an equal proportion, with each month recording 8.3%

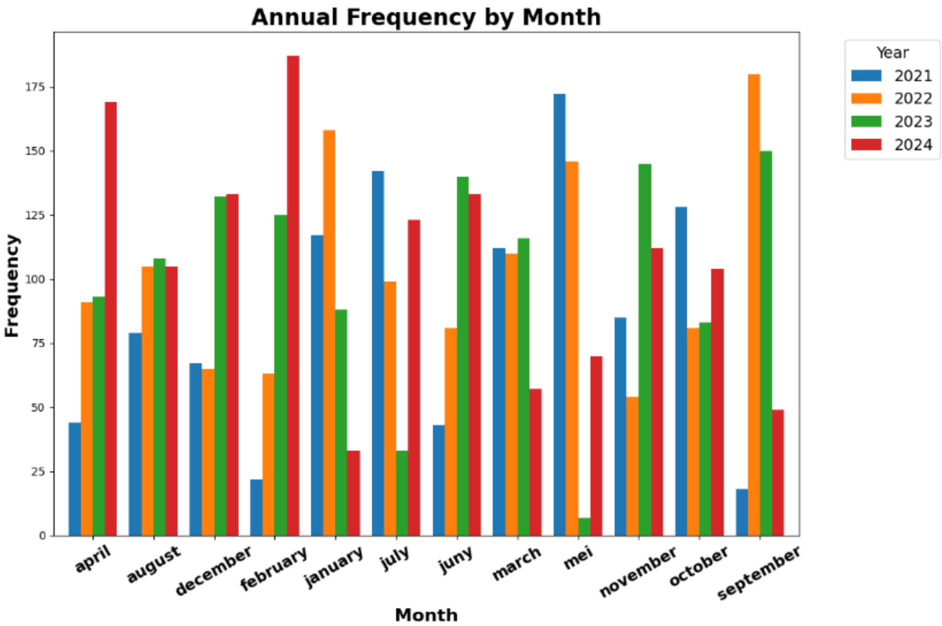


Figure 7. Annual Frequency by Month

Based on the analysis of the graph, July, March, and May were identified as the periods with the highest frequency of Kamtibmas incidents within the jurisdiction of the Bireuen District Police during the 2021 to 2024 period. July demonstrated consistency as a month with a high frequency of incidents, maintaining this trend until 2024, which highlights its significance in the annual cycle of Kamtibmas disturbances. On the other hand, March experienced a significant decline in the last year of the analysis period, suggesting a potential shift in the social dynamics influencing Kamtibmas incidents during this month. May and June recorded a surge in incidents in 2021, followed by a decline, but incidents increased again in 2024, showing a fluctuating yet notable pattern of disturbances in these months. Several other months, such as April, August, and December, also exhibited an upward trend in the last two years, pointing to a potential seasonal or periodic rise in incidents during these months. In contrast, October experienced sharp fluctuations, indicating irregular patterns or specific incidents that may have contributed to these spikes. Meanwhile, January, February, and November tended to maintain more stable frequencies of incidents, reflecting a more consistent pattern of Kamtibmas incidents. This overall trend suggests the existence of seasonal dynamics that influence the intensity and occurrence of Kamtibmas incidents, which must be taken into account when planning princidentive measures and anticipating future security challenges.

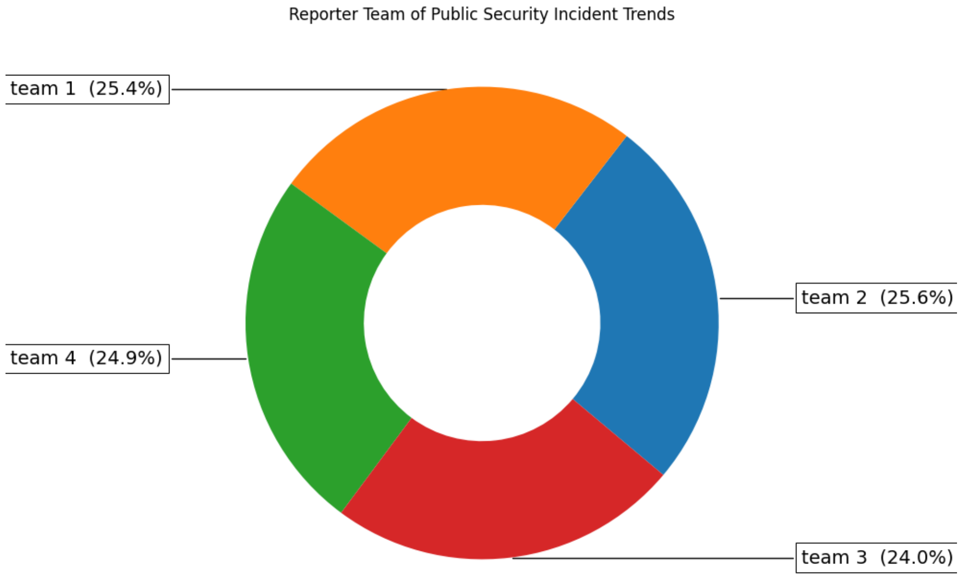


Figure 8. Reporter Team of Public Security Incident Trends

Based on the data in the graph above, it can be observed that the distribution of teams handling Kamtibmas incidents is relatively even, with each team playing a significant role in addressing security disturbances within the jurisdiction. Specifically, Team 1 holds the largest proportion, accounting for 25.4% of the total incidents, followed closely by Team 2 at 25.6%. Team 4 is slightly behind with 24.9%, while Team 3 handles 24.0% of the cases. This distribution indicates that the contributions of each team in managing Kamtibmas incidents are nearly balanced, suggesting effective coordination and resource allocation among the teams in addressing security concerns within the region. Such an even spread of responsibilities is crucial for maintaining an organized and efficient approach to handling public security disturbances.

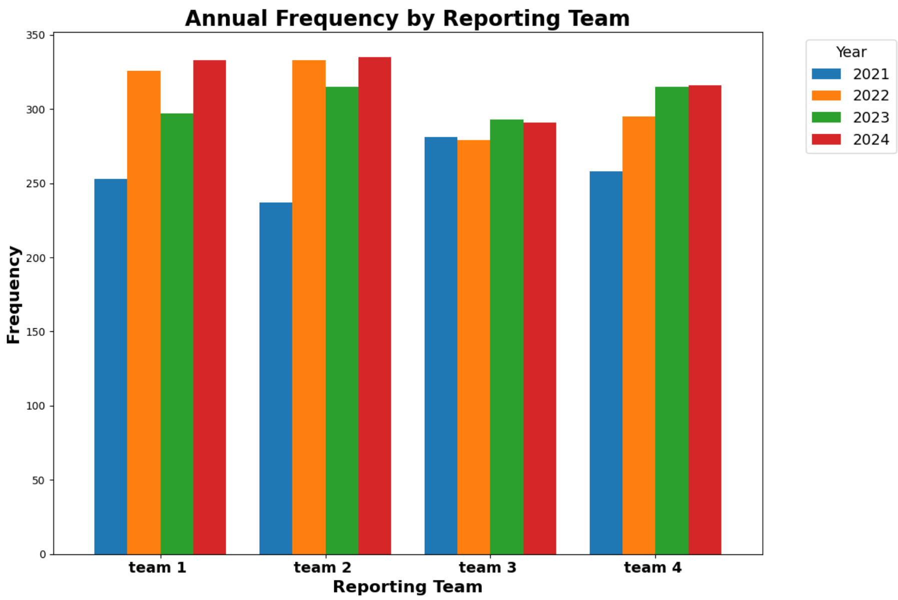


Figure 9. Annual Frequency by Reporting Team

Based on the graph shown, the number of Kamtibmas incident reporters increased steadily from 2021 to 2024, with notable spikes observed in Team 1 and Team 2, particularly in 2022 and 2024. These sharp increases suggest that these teams may have encountered more significant or complex security issues during these years, which prompted a higher frequency of reporting. On the other hand, Team 3 exhibited a trend that remained relatively stable, with only a slight increase in the number of reports, indicating a consistent level of activity without dramatic fluctuations. Meanwhile, Team 4 experienced a gradual rise in the number of incident reports, culminating in the highest reporting numbers in 2023 and 2024, which could indicate a growing involvement or a shift in focus towards more intensive security monitoring. Overall, 2024 saw the highest number of reports across all teams, reflecting not only an increase in the intensity of Kamtibmas incidents but also a higher level of engagement and participation from the teams in documenting and responding to security disturbances. This trend underscores the evolving nature of public security and the increasing capacity of the reporting teams in managing and addressing these incidents.

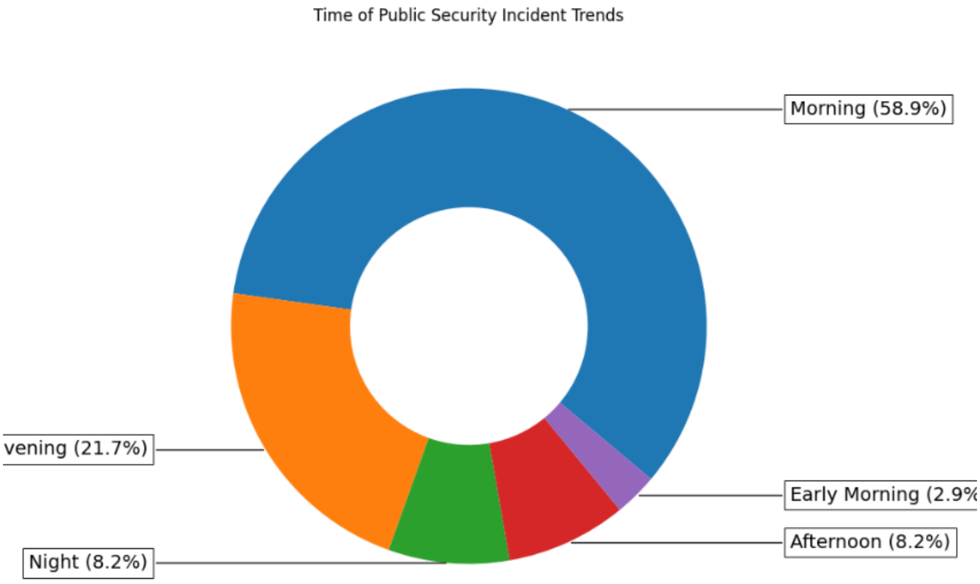


Figure 10. Time of Public Security Incident Trends

Based on the data in the graph above, it can be observed that the majority of Kamtibmas incidents occur in the morning, with a significant percentage of 58.9%. This suggests that the morning hours are the most active period for security disturbances, which could be attributed to various factors such as increased public activity, commuting, and other social dynamics that occur during this time. Following the morning, the incidence of Kamtibmas drops notably in the afternoon, accounting for 21.7%, and remains relatively low at night and during the day, with each period reporting 8.2% of incidents. The early morning hours report the fewest incidents, with only 2.9%. This distribution indicates that while Kamtibmas incidents can occur at any time of the day, mornings are by far the most common time for disturbances, which may point to the need for heightened security measures or princidentive strategies during these hours.

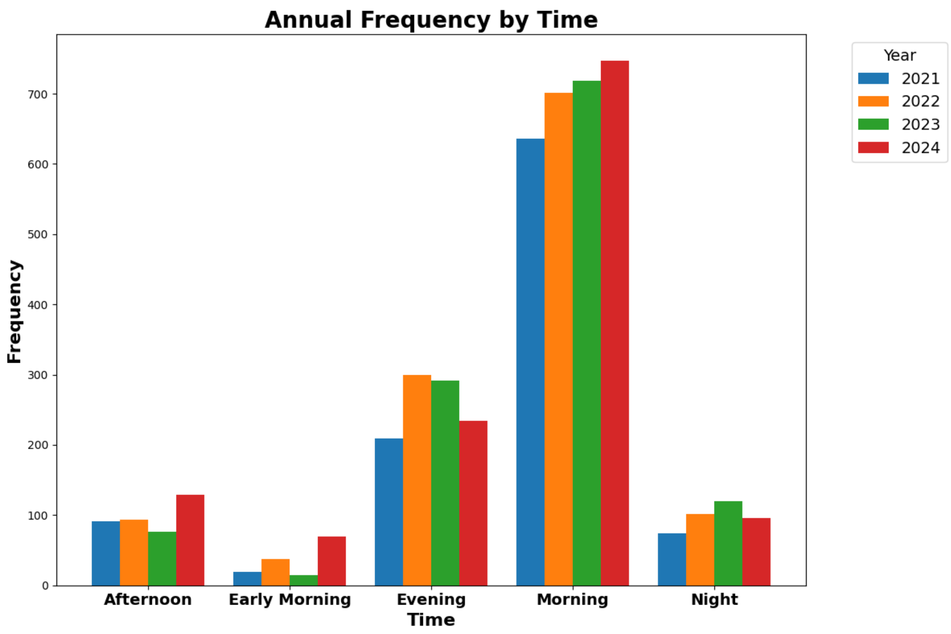


Figure11. Annual Frequency by Time

Based on the graph shown, the morning consistently recorded the highest frequency of Kamtibmas incidents from 2021 to 2024, with an upward trend that peaked in 2024, making it the most vulnerable period for disturbances. The afternoon also saw relatively high numbers of incidents in 2022 and 2023, though it experienced a noticeable decline in 2024. Meanwhile, the daylight period showed fluctuating trends but surged sharply in 2024, suggesting a possible shift in the timing of Kamtibmas incidents. Nighttime incidents tended to remain stable throughout the years, and early morning reports, while still the lowest, showed a significant spike in the final year of the analysis. This pattern highlights the importance of maintaining heightened vigilance and adapting to the evolving timing dynamics of Kamtibmas incidents.

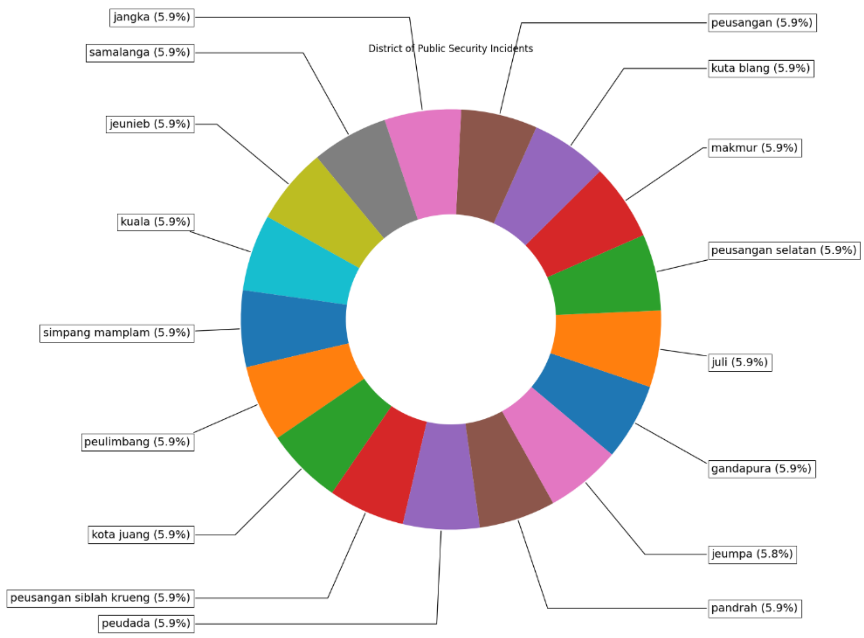


Figure 12. District of Public Security Incidents

Based on the graph shown in the figure above, the distribution of Kamtibmas incidents across the 17 sub-districts in Bireuen District is relatively even, with each sub-district contributing 5.9% of the total incidents, except for Jeumpa District, which recorded a slightly lower proportion of 5.8%.

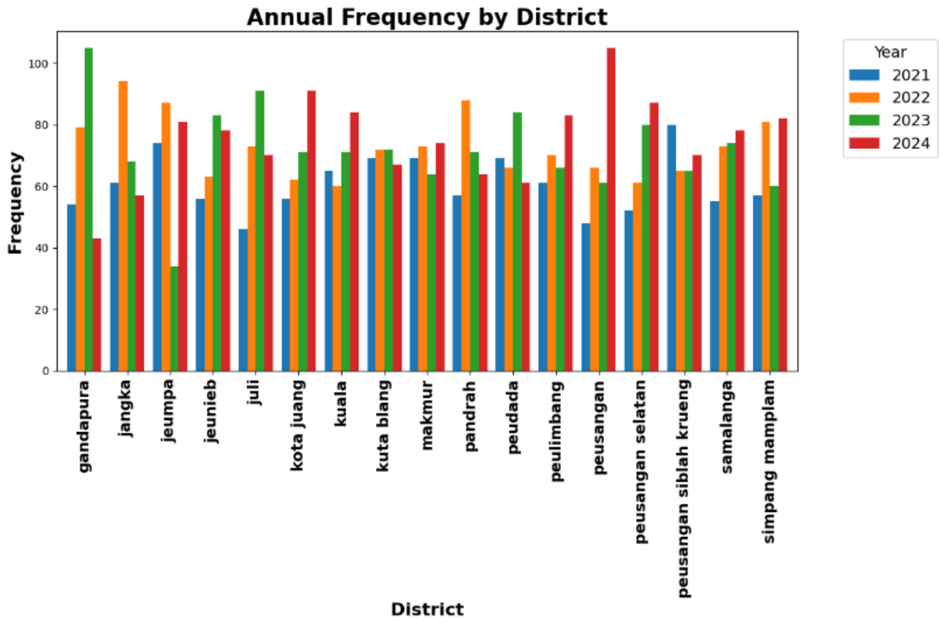


Figure 13. Annual Frequency by District

Figure 13 shows variations in the frequency of Kamtibmas incidents across each sub-district under the jurisdiction of the Bireuen Police from 2021 to 2024, revealing distinct patterns. Some sub-districts, such as Kuala, Kota Juang, and Kuta Blang, demonstrate relatively stable trends. In contrast, Pandrah and South Peusangan experienced significant spikes in incidents in 2023 and 2024, respectively. The Mamplam junction exhibited a consistent increase, suggesting a growing risk, while Gandapura experienced a sharp decline. Irregular fluctuations were also observed in sub-districts such as Peusangan, Peulimbang, Peudada, and Jeumpa. These findings highlight the importance of adapting Kamtibmas strategies to the specific dynamics and context of each region.

Discussion

Naïve Bayes Algorithm

In this study, the Gaussian Naïve Bayes algorithm was applied with a var_smoothing parameter of 0.5. This was done to ensure numerical stability and princident issues like zero division errors. Additionally, small noise was added to the feature data to improve the model's generalization abilities and reduce the risk of overfitting. For feature selection, the Mutual Information method was utilized with a value of k = 2. This method filters out the most relevant features in relation to the target variable, enhancing the efficiency and accuracy of the model. To evaluate the model's performance, macro-average metrics—accuracy, precision, recall, and F1-score—were used to ensure a fair assessment, particularly when dealing with unbalanced data.

Accuracy of Naive Bayes: 0.42

Classification Report:

	precision	recall	f1-score	support
ekonomi di daerah_1	0.36	0.67	0.47	153
ekonomi di daerah_2	0.41	0.72	0.52	183
ekonomi di daerah_3	0.00	0.00	0.00	142
ekonomi di daerah_4	0.00	0.00	0.00	100
keamanan di daerah_1	0.43	0.53	0.48	158
keamanan di daerah_2	0.33	0.72	0.45	141
keamanan di daerah_3	0.00	0.00	0.00	131
keamanan di daerah_4	1.00	0.04	0.07	110
politik di daerah_1	0.46	0.79	0.58	218
politik di daerah_2	0.40	0.82	0.54	204
politik di daerah_3	0.00	0.00	0.00	177
politik di daerah_4	0.00	0.00	0.00	123
sosial budaya di daerah_1	0.53	0.74	0.62	170
sosial budaya di daerah_2	0.50	0.72	0.59	161
sosial budaya di daerah_3	0.17	0.01	0.02	115
sosial budaya di daerah_4	0.00	0.00	0.00	93
accuracy			0.42	2379
macro avg	0.29	0.36	0.27	2379
weighted avg	0.31	0.42	0.32	2379

Figure 14. Naïve Bayes 50:50 Ratio Classification Report

The experiments conducted with a 50% training and 50% testing data split resulted in an accuracy of 0.42, with a precision of 0.29 and a recall of 0.36. These results indicate that the model struggles with accurately recognizing positive instances. The F1-score of 0.27 highlights the imbalance between precision and recall, further suggesting that the model's overall performance in classification is suboptimal in this data-sharing scenario.

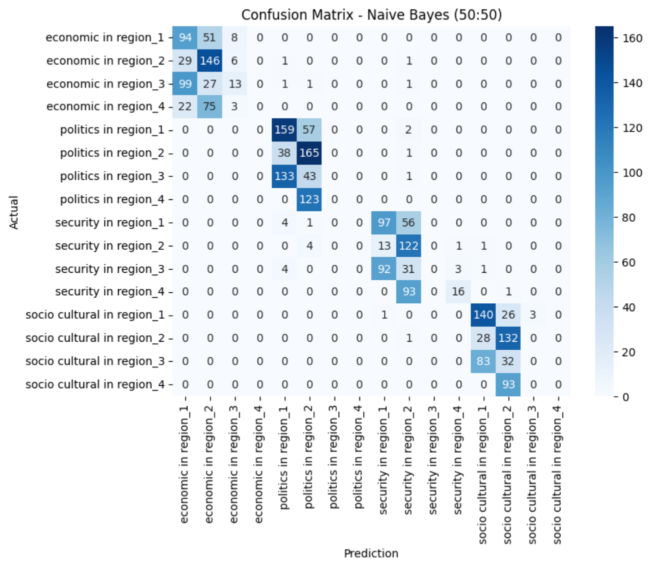


Figure 15. Confusion Matrix Ratio 50:50 Naïve Bayes

The Naïve Bayes model with a 50:50 data ratio yielded the best classification results in political classes for region_1 and region_2, as well as security in region_4 and socio-cultural issues in some regions, despite an overall accuracy of only 42%. However, significant prediction errors were observed, particularly in the economy and security categories. For instance, the economy in region_1 was frequently misclassified as the economy in region_2, and security in region_1 was often mistakenly identified as security in region_2. These errors suggest overlapping patterns between classes, supporting findings by Heydarian et al. (2022) that a multi-label confusion matrix-based approach can help reduce error rates in categories with similar patterns.

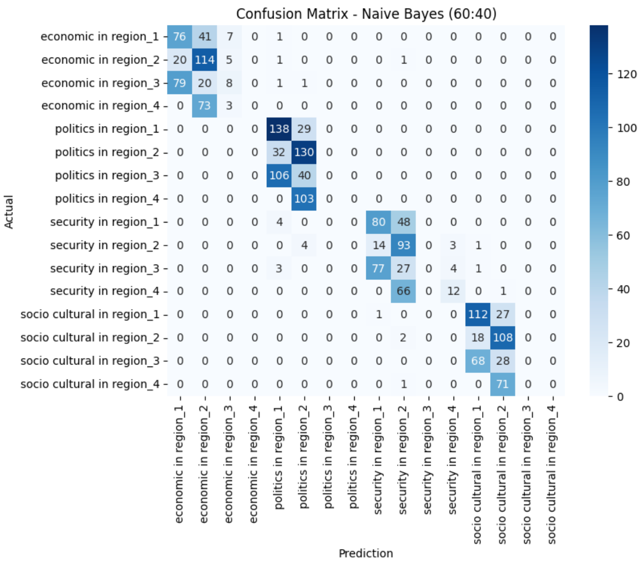
Accuracy of Naive Bayes: 0.44

Classification Report:

	precision	recall	f1-score	support
ekonomi di daerah_1	0.39	0.69	0.50	125
ekonomi di daerah_2	0.41	0.73	0.53	141
ekonomi di daerah_3	0.00	0.00	0.00	109
ekonomi di daerah_4	0.00	0.00	0.00	76
keamanan di daerah_1	0.44	0.55	0.49	132
keamanan di daerah_2	0.38	0.75	0.51	115
keamanan di daerah_3	0.00	0.00	0.00	112
keamanan di daerah_4	0.82	0.23	0.36	79
politik di daerah_1	0.46	0.82	0.59	167
politik di daerah_2	0.39	0.80	0.53	162
politik di daerah_3	0.00	0.00	0.00	146
politik di daerah_4	0.00	0.00	0.00	103
sosial budaya di daerah_1	0.59	0.79	0.67	140
sosial budaya di daerah_2	0.49	0.75	0.59	128
sosial budaya di daerah_3	0.00	0.00	0.00	96
sosial budaya di daerah_4	0.00	0.00	0.00	72
accuracy			0.44	1903
macro avg	0.27	0.38	0.30	1903
weighted avg	0.29	0.44	0.34	1903

Figure 16. Naïve Bayes 60:40 Ratio Classification Report

The experiment with a 60% training and 40% testing data split resulted in a model accuracy of 0.44, with a precision of 0.27 and a recall of 0.38. This indicates that the model is still not optimal in accurately identifying positive data. The F1-score value of 0.30 suggests that the balance between precision and recall is still low, meaning the model's performance in classifying Kamtibmas trends at this ratio has not shown a significant improvement compared to the previous ratio (50% train - 50% test). Although there is a slight increase in accuracy, errors in predicting certain categories still indicate that the model is not yet fully effective in handling imbalanced data or categories with similar patterns.



Gambar 17. Confusion Matrix Ratio 60:40 Naïve Bayes

The Naïve Bayes model with a 60:40 data ratio managed to classify 1903 Kamtibmas incidents and showed the best performance of the political class in region_2, region_1, and region_4 with accurate predictions of 130, 137, and 103 incidents, respectively. However, there are still significant misclassifications, particularly in the economy and security classes, such as the economy in region_1 and region_3 which is often misclassified as the economy in region_2, and security in region_1 and region_3 which are mistakenly predicted to be other classes. This suggests the similarity of features between subcategories and data imbalances, which according to (Grandini et al., 2020) are often the cause of errors in multi-class classification.

Accuracy of Naive Bayes: 0.43

Classification Report:

	precision	recall	f1-score	support
ekonomi di daerah_1	0.34	0.50	0.40	103
ekonomi di daerah_2	0.35	0.70	0.47	105
ekonomi di daerah_3	0.00	0.00	0.00	82
ekonomi di daerah_4	0.00	0.00	0.00	60
keamanan di daerah_1	0.42	0.55	0.48	94
keamanan di daerah_2	0.35	0.71	0.47	83
keamanan di daerah_3	0.00	0.00	0.00	78
keamanan di daerah_4	1.00	0.15	0.26	59
politik di daerah_1	0.47	0.88	0.61	122
politik di daerah_2	0.42	0.80	0.55	128
politik di daerah_3	0.00	0.00	0.00	99
politik di daerah_4	0.00	0.00	0.00	74
sosial budaya di daerah_1	0.57	0.77	0.66	111
sosial budaya di daerah_2	0.50	0.72	0.59	101
sosial budaya di daerah_3	0.00	0.00	0.00	72
sosial budaya di daerah_4	0.00	0.00	0.00	57
accuracy			0.43	1428
macro avg	0.28	0.36	0.28	1428
weighted avg	0.30	0.43	0.33	1428

Figure 18. Naïve Bayes 70:30 Ratio Classification Report

Experiments with a 70% training data and 30% test data ratio yielded an accuracy of 0.43, with a precision of 0.28 and a recall of 0.36, suggesting that the model was only able to correctly identify a small portion of the positive instances. The F1-score of 0.28 reflects a low balance between precision and recall, indicating that the Naïve Bayes model at this ratio is still not optimal for accurately classifying the categories of Kamtibmas incidents.

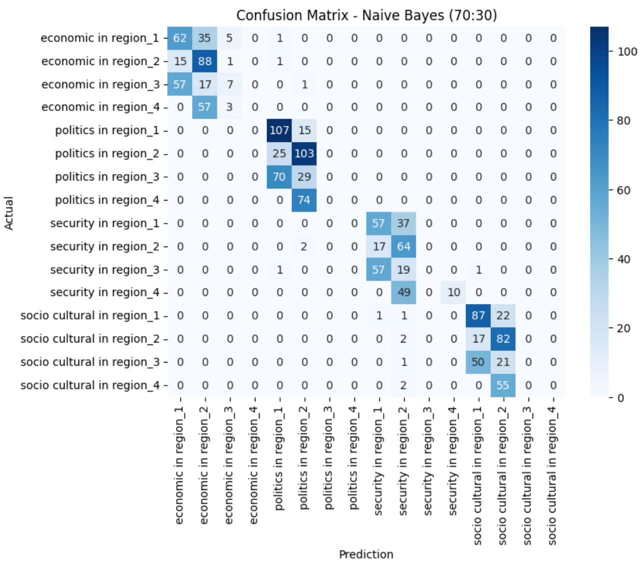


Figure 19. Confusion Matrix Ratio 70:30 Naïve Bayes

The Naïve Bayes model with a 70:30 data ratio successfully classified 1,428 Kamtibmas incidents, demonstrating relatively stable performance, particularly in the political category for region_2 and region_1, as well as the socio-cultural category in region_1. Other classes, such as politics in region_4 and economy in region_3, also yielded promising results, reflecting the model's improved pattern recognition when trained on a larger dataset—even though the overall accuracy remained at 43%. However, notable misclassifications persist, such as the frequent confusion between economy in region_1 and region_2, and inaccuracies in the security category across several regions. These issues suggest that the model still struggles to differentiate between subcategory characteristics. This aligns with the findings of Muttaqin, Wahyu Wijaya Widiyanto et al. (2023), which highlight the usefulness of a confusion matrix in evaluating and identifying ambiguities in classification performance, thereby indicating the need for further optimization.

Accuracy of Naive Bayes: 0.42

Classification Report:

	precision	recall	f1-score	support
ekonomi di daerah_1	0.34	0.49	0.40	73
ekonomi di daerah_2	0.35	0.69	0.46	70
ekonomi di daerah_3	0.00	0.00	0.00	61
ekonomi di daerah_4	0.00	0.00	0.00	39
keamanan di daerah_1	0.45	0.68	0.54	57
keamanan di daerah_2	0.39	0.72	0.51	57
keamanan di daerah_3	0.00	0.00	0.00	48
keamanan di daerah_4	1.00	0.16	0.28	43
politik di daerah_1	0.47	0.84	0.60	81
politik di daerah_2	0.38	0.82	0.52	78
politik di daerah_3	0.00	0.00	0.00	64
politik di daerah_4	0.00	0.00	0.00	57
sosial budaya di daerah_1	0.55	0.74	0.63	70
sosial budaya di daerah_2	0.48	0.71	0.57	68
sosial budaya di daerah_3	0.00	0.00	0.00	47
sosial budaya di daerah_4	0.00	0.00	0.00	39
accuracy			0.42	952
macro avg	0.28	0.37	0.28	952
weighted avg	0.29	0.42	0.32	952

Figure 20. Report Classification Ratio 80:20 Naïve Bayes

Experiments with an 80% training and 20% test data split yielded an accuracy of 0.42, with a precision of 0.28 and a recall of 0.37. These results indicate that the model is still not optimal in accurately identifying positive cases. The F1-score of 0.28 reflects a low balance between precision and recall, suggesting that the overall performance of the model in classifying Kamtibmas trends at this ratio remains limited.

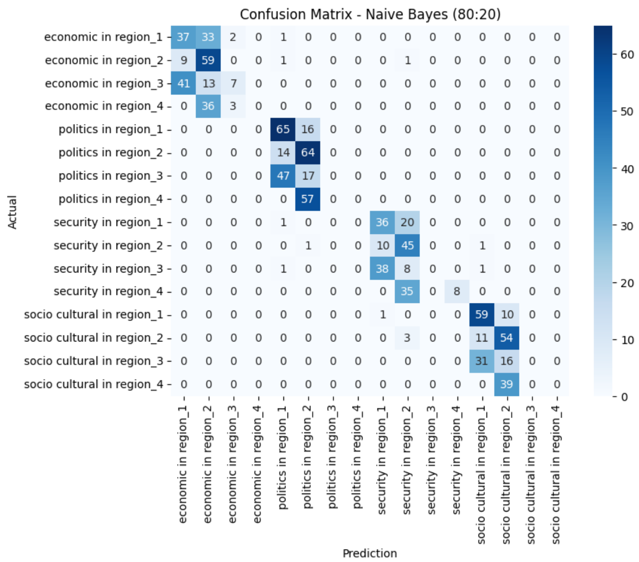


Figure 21. Confusion Matrix Ratio 80:20 Naïve Bayes

In experiments with an 80% training and 20% test data ratio, the Naïve Bayes model successfully classified 952 Kamtibmas incidents, showing fairly consistent results compared to previous ratios, although a decrease in precision was observed due to the smaller test set.

The model demonstrated its best performance in the political category within region_1, region_2, and region_4, as well as in the socio-cultural category in region_2. However, significant misclassifications still occurred, particularly in the economic and security categories that share similar features—for instance, economic incidents in region_1 and region_3 were often misclassified as belonging to region_2. This indicates that the model struggles to distinguish overlapping patterns between subcategories. Therefore, further optimization is necessary, such as enhancing feature quality or applying more advanced classification techniques. As suggested by Heydarian et al. (2022), a hierarchical classification approach could help reduce misclassification in categories with similar characteristics.

Accuracy of Naive Bayes: 0.42

Classification Report:

	precision	recall	f1-score	support
ekonomi di daerah_1	0.30	0.51	0.38	37
ekonomi di daerah_2	0.39	0.62	0.48	39
ekonomi di daerah_3	0.00	0.00	0.00	33
ekonomi di daerah_4	0.00	0.00	0.00	13
keamanan di daerah_1	0.45	0.62	0.52	29
keamanan di daerah_2	0.35	0.68	0.46	28
keamanan di daerah_3	1.00	0.04	0.08	24
keamanan di daerah_4	1.00	0.17	0.29	24
politik di daerah_1	0.52	0.88	0.66	43
politik di daerah_2	0.36	0.84	0.51	37
politik di daerah_3	0.00	0.00	0.00	31
politik di daerah_4	0.00	0.00	0.00	31
sosial budaya di daerah_1	0.46	0.72	0.56	29
sosial budaya di daerah_2	0.55	0.67	0.60	39
sosial budaya di daerah_3	0.00	0.00	0.00	20
sosial budaya di daerah_4	0.00	0.00	0.00	19
accuracy			0.42	476
macro avg	0.34	0.36	0.28	476
weighted avg	0.35	0.42	0.33	476

Figure 22. Naïve Bayes 90:10 Ratio Classification Report

Experiments using a 90% training and 10% test data split resulted in a model accuracy of 0.42, with a precision of 0.34 and recall of 0.36. These metrics suggest that the model struggled to identify positive cases accurately. The F1-score of 0.28 indicates poor balance between precision and recall, highlighting the Naïve Bayes model's difficulty in detecting positive classes effectively. This is particularly problematic in scenarios where accurately predicting positive cases is crucial, such as Kamtibmas trend analysis. The small test set and potential feature correlations may also have contributed to the model's limited performance. Further optimization or alternative algorithms may be needed to improve classification accuracy.

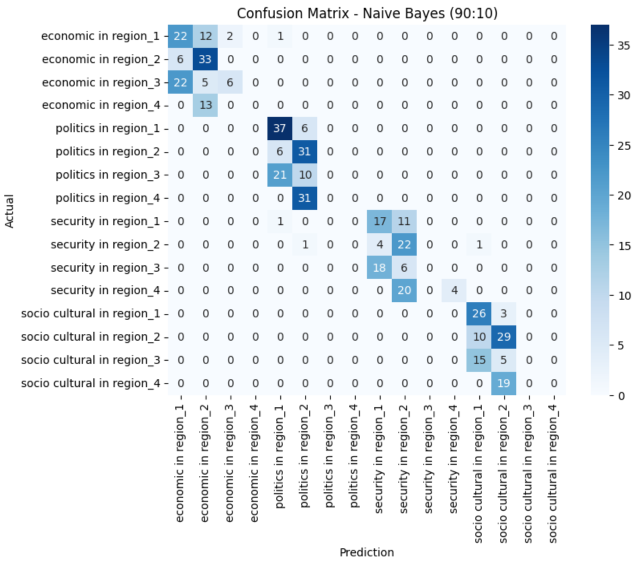


Figure 23. Confusion Matrix Ratio 90:10 Naïve Bayes

The confusion matrix at a 90:10 data ratio shows that the Naïve Bayes model was able to classify several categories fairly well, such as political incidents in region_1 and region_4, as well as economic incidents in region_3, even though the test set consisted of only 476 cases. However, the limited test data affected the model's accuracy and precision in several other classes—particularly security in region_1 and economics in region_2—which experienced frequent misclassifications due to feature similarities among subcategories. This highlights the model's continued difficulty in distinguishing between closely related classes. As noted by Heydarian et al. (2022), utilizing a multi-label confusion matrix and a hierarchical classification approach can help reduce errors in categories with overlapping patterns.

Random Forest Algorithm

The implementation of the Random Forest model was conducted using 100 decision trees with a maximum depth of 10 to prevent overfitting, along with a random_state parameter set to 42 to ensure result consistency. This model was selected due to its ability to aggregate multiple trees for more accurate and stable predictions, as well as its capability to manage non-linear data with complex feature interactions. In experiments using a 50:50 train-test data split, the model demonstrated strong performance, achieving an accuracy of 0.95, precision of 0.96, recall of 0.95, and an F1-score of 0.95. These metrics indicate a significantly better classification balance compared to the previous model.

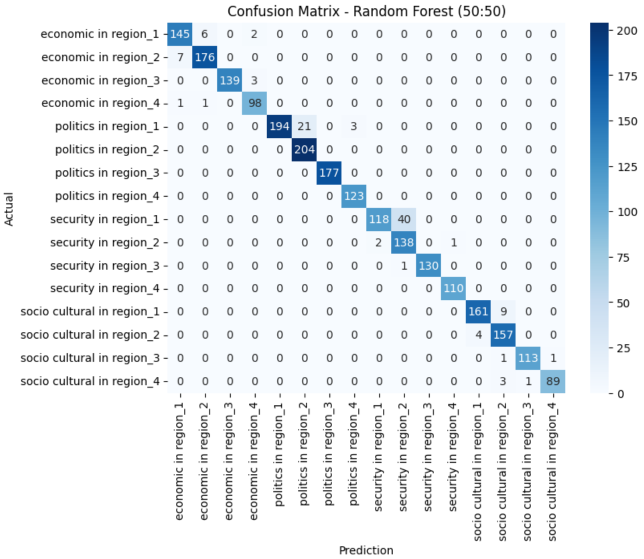


Figure 24. Confusion Matrix Ratio 50:50 Random Forest

The confusion matrix from the Random Forest model using a 50:50 data split demonstrated outstanding performance in classifying 2,379 Kamtibmas incidents. The model achieved high accuracy in predicting political categories in region_2 and region_1, as well as economic incidents in region_2. Other classes, including socio-cultural and security incidents in several regions, were also accurately classified, highlighting the model's strong capability to differentiate between class features effectively. Although some misclassifications still occurred—particularly in the security and political categories within region_1—the overall proportion of errors was relatively low. As noted by Muttaqin, Wahyu Wijaya Widiyanto et al. (2023), the confusion matrix remains a critical tool for evaluating model performance and pinpointing areas that require further optimization.

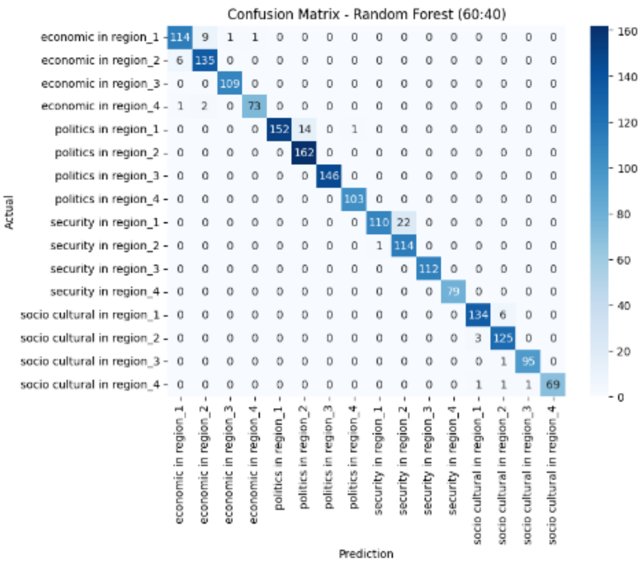


Figure 25. Confusion Matrix Ratio 60:40 Random Forest

In experiments using a 60% training data and 40% test data ratio, the Random Forest model demonstrated exceptional performance, achieving an accuracy of 0.96, with precision and recall both at 0.97, and an excellent f1-score, indicating a well-balanced classification. Out of a total of 1,903 Kamtibmas incidents, the model effectively classified most categories, particularly political events in region_2 and region_1, as well as socio-cultural and economic incidents across several regions. The confusion matrix revealed that although there were a few misclassifications—specifically in the security and political categories in region_1—their impact on the model's overall performance was minimal. These results suggest that Random Forest is a highly reliable model for detecting Kamtibmas patterns, especially when trained with a larger dataset.

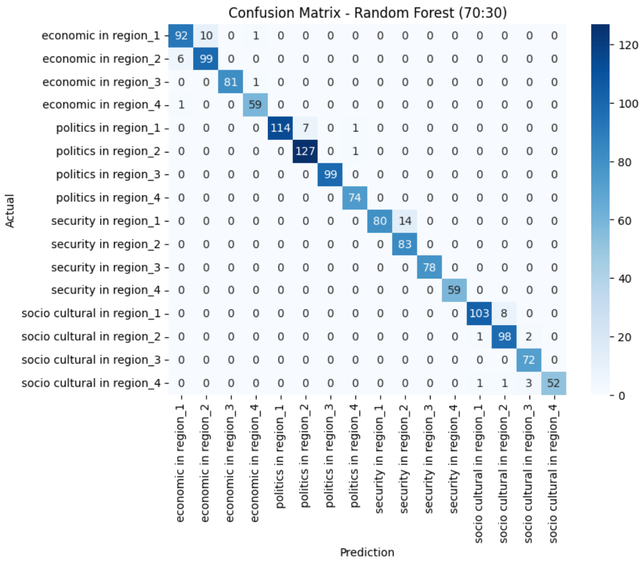


Figure 26. Confusion Matrix Ratio 70:30 Random Forest

In experiments with a 70% training data and 30% test data ratio, the Random Forest model achieved an accuracy of 0.96, with precision and recall of 0.97 and 0.96, respectively, and an f1-score of 0.96, reflecting its excellent performance. Out of 1,428 Kamtibmas incidents, the model successfully classified most categories, particularly political events in region_2, politics in region_1, and economics in region_2. However, some misclassifications were still observed, particularly in categories with similar patterns, such as security and politics in region_3, and minor classes that tended to spread to other categories. These results indicate the need for improvements, including data balancing and feature selection, to enhance prediction accuracy, particularly for more complex categories.

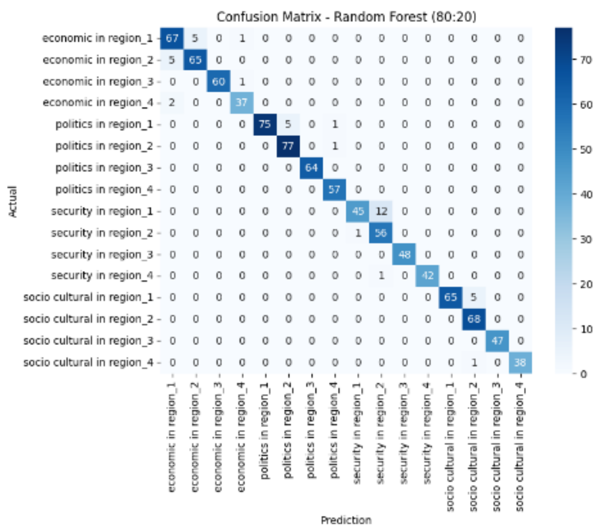


Figure 27. Confusion Matrix Ratio 80:20 Random Forest

In experiments with an 80% training data and 20% test data ratio, the Random Forest model achieved an accuracy of 0.97, with consistently high precision, recall, and f1-score of 0.97, indicating excellent and balanced classification performance. Out of a total of 952 Kamtibmas incidents, the model accurately classified most categories, particularly political and socio-cultural classes in several regions. However, minor errors were observed in economic, security, and socio-cultural subclasses with similar characteristics. For instance, the economy in region_2 was sometimes misclassified as the economy in region_1. While the overall performance is excellent, these misclassifications underscore the need for further optimization to improve the accuracy of classifying overlapping subcategories, as confirmed by (Muttaqin, Wahyu Wijaya Widiyanto et al., 2023) through evaluation using a confusion matrix.

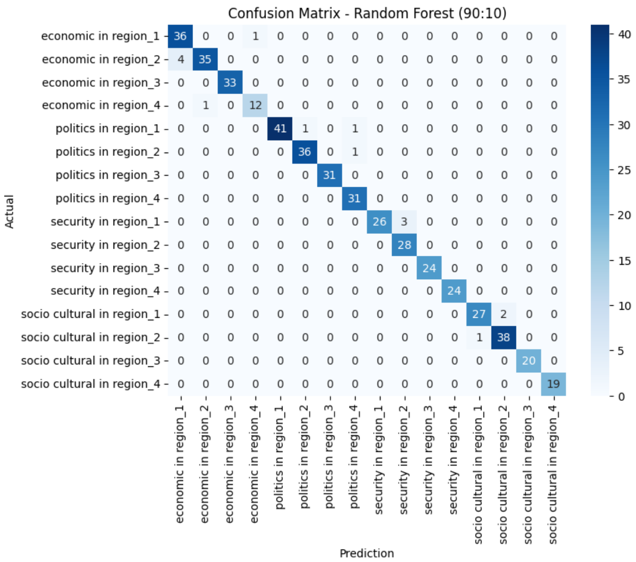


Figure 28. Confusion Matrix Ratio 90:10 Random Forest

In experiments with a 90% training data and 10% test data ratio, the Random Forest model demonstrated very high performance, achieving accuracy, precision, recall, and f1-score of 0.98 each, despite the relatively small amount of test data. The model was able to classify categories such as politics in region_1, socio-culture in region_2, and security and economics in some regions with great accuracy, indicating consistency in recognizing patterns of Kamtibmas incidents. However, minor errors still occurred in subcategories with similar characteristics, such as economics in region_2 and security in region_1. This aligns with the findings of Heydarian et al. (2022), who suggest that the multi-label confusion matrix can help reduce errors in classes with similar features. Overall, this model shows great potential as a predictive tool for the Bireuen Police in both the prevention and handling of Kamtibmas incidents.

Machine Learning Evaluation Model

The evaluation of the machine learning model was conducted by comparing the macro averages of accuracy, precision, recall, and f1-score values across different data-sharing scenarios for both the Naïve Bayes and Random Forest algorithms. The results of this comparison are summarized in Table 1. Overall, the Random Forest algorithm demonstrates significantly superior performance compared to Naïve Bayes, particularly in terms of accuracy and the consistency of classifying Kamtibmas trend data within the jurisdiction of the Bireuen District Police.

Table 1. Summary of Machine Learning Algorithm Test Results

Algoritma	Ratio	Accuracy	Precision	Recall	F1-score
NB	50 : 50	0.42	0.29	0.36	0.27
	60 : 40	0.44	0.27	0.38	0.30
	70 : 30	0.43	0.28	0.36	0.28
	80 : 20	0.42	0.28	0.37	0.28
	90 : 10	0.42	0.34	0.36	0.28
RF	50 : 50	0.95	0.96	0.95	0.95
	60 : 40	0.96	0.97	0.97	0.97
	70 : 30	0.96	0.97	0.96	0.96
	80 : 20	0.97	0.97	0.97	0.97
	90 : 10	**0.98**	**0.98**	**0.98**	**0.98**

Based on the test results, the Random Forest algorithm with a 90:10 training and test data ratio demonstrated the best performance in predicting Kamtibmas trends, achieving an accuracy, precision, recall, and f1-score of 0.98 each. At lower ratios, such as 80:20 and 50:50, the performance remains high (ranging from 0.95 to 0.97), showcasing Random Forest's consistency and reliability. In comparison, the Naïve Bayes algorithm underperforms, with accuracy ranging between 0.42 and 0.44 and low f1-scores, indicating its limitations in handling the Kamtibmas data. The Random Forest algorithm's success in handling complex datasets is further supported by previous research (Buani, 2024; Di Martino et al., 2024), which highlights its effectiveness in processing diverse input variables and maintaining stable prediction results. Given its robust performance, Random Forest is highly suitable as a data-driven decision-making tool for mapping and anticipating Kamtibmas trends in the jurisdiction of the Bireuen District Police.

CONCLUSION

This study successfully applied the Naïve Bayes and Random Forest algorithms to predict Kamtibmas trends in the jurisdiction of the Bireuen Police using historical data from daily intelligence reports. The process included stages such as data collection, preprocessing, splitting the data into training and test sets, and applying both algorithms for incident classification. Among the two algorithms tested, Random Forest demonstrated superior performance, achieving accuracy, precision, recall, and F1-score values of 98% at a 90:10 training and test data ratio. In contrast, Naïve Bayes proved to be less effective due to its limitations in handling data with high correlation between variables, resulting in suboptimal performance.

REFERENCES

Akbar, S. M. (2024). Integritas Kepolisian Region Bengkulu dalam Pengawasan Pemilihan Kepala Region Prespektif Fiqh Siyasah (Studi Kasus pada Polisi Region Bengkulu). Skripsi, 37–48.

Buani, D. C. P. (2024). Deteksi Dini Penyakit Diabetes dengan Menggunakan Algoritma Random Forest. EVOLUSI : Jurnal Sains Dan Manajemen, 12(1), 1–8. https://doi.org/10.31294/evolusi.v12i1.21005

Di Martino, S., Landolfi, E., Mazzocca, N., Rocco di Torrepadula, F., & Starace, L. L. L. (2024). A visual-based toolkit to support mobility data analytics. Expert Systems with

Applications, 238(PC), 121949. https://doi.org/10.1016/j.eswa.2023.121949
Fakhri, J., Sunge, A. S., & Turmudi zy, A. (2023). Perancangan Classification Algoritma Naive Bayes Pada Data Pemilihan Jurusan Siswa. JTT (Jurnal Teknologi Terpadu), 11(2), 260–269. https://doi.org/10.32487/jtt.v11i2.1823
Febrera, G., & Prianggono, J. (2024). Classification of Traffic Violations Using the Naïve Bayes Algorithm at Padang City Police. Penelitian Ilmu Komputer, Sistem Embedded and Logic, 12(225), 481–492. https://doi.org/10.33558/piksel.v12i2.10018
Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for Multi-Class Classification: an Overview. White Paper, 2(3), 1–17. http://arxiv.org/abs/2008.05756
Hadmanto, A., & Prianggono, J. (2024). Prediction of Theft with Machine Learning Technology at Police Station. Building of Informatics, Technology and Science (BITS), 6(1), 550–558. https://doi.org/10.47065/bits.v6i1.5107
Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. IEEE Access, 10, 19083–19095. https://doi.org/10.1109/ACCESS.2022.3151048
Lianasari, M., & Ahmadi, S. (2022). Pengaruh Kompetensi dan Lingkungan Kerja Terhadap Kinerja dengan Motivasi Kerja sebagai Variabel Intervening. Fokus Bisnis Media Pengkajian Manajemen Dan Akuntansi, 21(1), 43–59. https://doi.org/10.32639/fokbis.v21i1.106
May, S., Isafiade, O., & Ajayi, O. (2021). An Enhanced Naïve Bayes Model for Crime Prediction using Recursive Feature Elimination. ACM International Conference Proceeding Series, 24(26), 580–586. https://doi.org/10.1145/3488933.3488973
Muttaqin, Wahyu Wijaya Widiyanto, M. M., Green Ferry Mandias, Stenly Richard Pungus, A. W., Wiranti Kusuma Hapsari, S. A. H., Aslam Fatkhudin, Pasnur, E. F. B., & Mochammad Anshori, Suryani, N. S. (2023). Pengenalan Data Mining (Issue July).
Purwatiningsih, B., & Polri, W. I. S. S. L. (2023). Optimalisasi Penanggulangan Konflik Sosial Guna Mendukung Pemilu Damai 2024 Dalam Rangka Memelihara Kamtibmas. Dalam SANYATA.
Raja Rizki Alanta Nasution, & Relita Buaton. (2024). Penerapan Metode Naive Bayes dalam Menentukan Diagnosa Kerusakan pada Smartphone. Indonesian Journal of Education And Computer Science, 2(1), 24–33. https://doi.org/10.60076/indotech.v2i1.383
Rifai, M. F., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). Petir, 12(2), 131–144. https://doi.org/10.33322/petir.v12i2.471
Rosmaya, I. N. A., Kirsten, O. N., & Putri, R. S. (2022). Peningkatan Keamanan dan Ketertiban Masyarakat (Kamtibmas) Dusun Nogosari. vol. 4(02), hlm. 1313–1322, Jurnal Abdi Bhayangkara
Sugiyono. (2021). Metode Penelitian Kuantitaif, Kualitatif, R&D. Alfabeta.
Supriyadi, R., Gata, W., Maulidah, N., & Fauzi, A. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis, 13(2), 67–75. https://doi.org/10.51903/e-bisnis.v13i2.247
Tampubolon, C. N. (2023). Analisis Yuridis Peran Intelkam Polda Kepri Dalam Mengidentifikasi Ancaman Tantangan Hambatan Gangguan Kamtibmas Guna Menegakkan Kepastian Hukum. UNISSULA Institutional Repository.
Wibowo, A. H., & Oesman, T. I. (2020). The comparative analysis on the accuracy of k-NN, Naive Bayes, and Decision Tree Algorithms in predicting crimes and criminal actions in Sleman Regency. Journal of Physics: Conference Series, 1450(1). https://doi.org/10.1088/1742-6596/1450/1/012076
Widya Bhakti Dira, J. P. (2024). Analisis perbandingan algoritma ann dan knn dalam prediksi tindak pidana pencurian kendaraan bermotor di polresta malang kota. Prosiding Seminar Nasional Sains Dan Teknologi Seri, 1(2), 440–460.

Zaidi, N., Mustapha, A., Mostafa, S., & Razali, M. N. (2020). A Classification Approach for Crime Prediction (pp. 68–78). https://doi.org/10.1007/978-3-030-38752-5_6