Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Exploration Of Violence Case Patterns In Jayapura Police Site Using Random Forest And K-Means

Widya Kurniawan[1], Eko Prasetio Widhi[2], Dewi Ikhlas Sari[3*]

[1,2,3] University of Darussalam Gontor;Jl. Raya Siman Km. 6, Siman, Ponorogo, East Java 63471; Tel/Fax: (0352) 483762
dewiikhlassari22@student.cs.unida.gontor.ac.id[3]*

Abstract

Crime is an act that endangers individuals and society, with significant social consequences, including psychological trauma and economic losses. In Jayapura City, violence cases have shown a notable increase from 133 cases in 2020 to 233 cases in 2024, with "common assault" and "mob violence" being the most dominant types. The primary issue in crime data management at the Jayapura Police Department lies in its manual and non-integrated system, hindering effective analysis and decision-making. This study applies data mining techniques using the Random Forest algorithm to classify case resolution status, K-Means Clustering to group cases by characteristics, and Holt-Winters Exponential Smoothing to forecast trends through 2025. The Random Forest model achieved 91% accuracy in classification, while clustering successfully identified three priority clusters. Forecasting results indicate a continued upward trend in violence cases in the upcoming year. These findings provide a strong foundation for data-driven policy formulation and enhance the efficiency of violence case resolution. With this comprehensive approach, it is expected that decision-making by law enforcement and stakeholders will become more responsive, targeted, and effective.
Keywords: *Crime; Violence; Data Mining; Random Forest; K-Means; Forecasting; Jayapura City*

1. Introduction

Crime is an act that endangers individuals and society, and has negative social impacts such as causing trauma. From a legal perspective, crime is a violation of the law that can be subject to criminal sanctions or fines. Examples include murder, robbery, theft, rape, and tax evasion[1].Acts of violence include various forms, such as physical violence, domestic violence, sexual violence, and violence against officers. Violence can cause physical and psychological suffering, and has the potential to cause trauma, psychological impacts, and even death for individuals and groups in society[2].

In Jayapura City, violence is a major issue in crime trends. As the center of social, governmental, and economic activities in Papua, the city has experienced a significant spike in cases of violence—from 133 cases in 2020 to 233 cases in 2024, an increase of around 73.19%. The most dominant types are ordinary assault and gang violence, which are categorized as Criminal Incidents (CI). This data also includes other metrics such as Case Total (CT), Case Completed (CC), and legal status such as SP3, ADR, and P21.
Jayapura Police, as a law enforcement under the Papua Police, faces major challenges in managing crime data. The recording system still uses Microsoft Excel manually, which is not integrated between units and is prone to input errors (human error). This makes it difficult to access data, slows down analysis, and hinders fast and accurate decision-making in handling the increasing cases of violence[3].
This study aims to identify patterns and factors that influence the occurrence of violence in Jayapura Regency during the 2020–2024 period. The approach used involves the Random Forest algorithm for classification and K-Means Clustering for grouping violence patterns. In addition, this study also seeks to develop a predictive model to estimate the risk of violence based on available historical data. The results of this analysis are expected to be used to provide data-based recommendations to the police and local governments as a basis for formulating more effective preventive policies in dealing with violence.

This study contributes to the development of the application of Data Mining techniques in the realm of crime analysis, especially by integrating two different algorithms in one analysis framework. On the practical side, this study is expected to improve the effectiveness of data-based decision-making in the Jayapura Police environment by providing accurate and easily interpreted information. Thus, the results of this study also have the potential to provide a positive impact on improving the security and comfort of the community in Jayapura City through more targeted and evidence-based interventions (evidence-based policy). Although the Random Forest and K-Means Clustering algorithms have been widely applied individually in crime studies for the purposes of data classification and grouping, most studies still use them separately without integration in a single integrated analytical framework. Amid the increasingly complex needs for data analysis in the big data era, the synergistic combination of these two algorithms, especially in a geographical context such as Jayapura Regency, has not been widely explored. In fact, Random Forest has advantages in high accuracy and resistance to overfitting on large datasets [4]. While K-Means is computationally efficient and produces clusters that are easy to interpret [5]. Therefore, this study aims to fill this gap by integrating the two algorithms to obtain a more comprehensive analysis of violence patterns and support data-based decision-making by authorities in the region.

2. Literature Review

Several previous studies have shown the success of applying the K-Means and Random Forest algorithms in complex data analysis. Chofidah et al. (2024) Comparing K-Means and Fuzzy Logic in grouping provincial crime indicators in Indonesia, with the results that K-Means is superior in terms of validity and stability of clustering[6]. Furthermore, Widya Kurniawan et al. (2024) Applying K-Means and PCA to group suicide cases in Central Java, producing high-quality clusters with a Silhouette Score reaching 84%[7]. Amelia Adhariani et al. (2024) Using K-Means in mapping crime vulnerabilities in Depok City, dividing areas into three risk levels with the CRISP-DM approach[8]. Fadil Danu Rahman (2024) Combining K-Means and Random Forest in analyzing weather data from BMKG. This combined model successfully grouped and predicted weather conditions with high accuracy, using the Elbow and Silhouette Score evaluations[9]. And finally Nurdiyanto Yusuf (2024) Used Random Forest Regression to predict national beef production with very accurate results (R^2 = 94.6%)[10].

2.1. Violence

In a narrow sense, violence refers to actions involving physical attacks, damage, or destruction, either against oneself or against objects that have the potential to belong to others. While in a broader sense, violence includes physical and psychological actions carried out by individuals or groups, either consciously or unconsciously, and can occur openly or secretly [11]. Acts of violence are acts carried out intentionally or due to negligence, which are included in violations of criminal law, carried out without justification or legitimate reasons, and are subject to sanctions by the state as a form of serious crime or minor violation of the law [12]. The Criminal Code states that anyone who jointly commits violence against people or objects in public can be sentenced to a maximum of five years and six months in prison [13].Based on Article 351 of the Criminal Code, abuse or acts of violence are defined as any form of act carried out intentionally with the aim of hurting or causing harm to someone's health, as well as acts intended to cause suffering to others. In another article it is stated that anyone who without the right forces another person to do, not do, or allow something through violence, unpleasant actions, or threats of violence or threats of unpleasant actions against another person, is considered to have violated the law[14].

2.2. Jurisdiction of Jayapura Police

The jurisdiction of Jayapura Police covers the entire area of Jayapura City, which consists of 5 sub-districts and 25 villages/sub-districts. This police station is responsible for maintaining security and order in the area. Jayapura City is the capital of Papua province with an area of 940 km or 940,000 Ha. Meanwhile, Jayapura City consists of 5 districts, 25 sub-districts, and 14 villages, namely North Jayapura, South Jayapura, Abepura, Heram, and Muara Tami. This city is located in the north bordering the Pacific Ocean, in the east directly bordering the neighboring country of Papua New Guinea, in the south bordering Kerom Regency, and in the west bordering Jayapura Regency. This makes the city very strategic and is one of the centers of trade and industry in the Eastern Indonesia region[15].

2.3. Data Mining

Data mining is a process intended to produce useful information from a large data set[16]. Data Mining is also an analysis process that aims to explore and assess large amounts of data in order to find patterns, trends, and useful information. This process begins with collecting data from various sources, such as databases, transactions, sensors, and social media, where the data collected must be relevant and of high quality so that the resulting analysis is accurate. After the data is collected, the next step is to clean the data to eliminate errors, duplications, and inconsistencies, so that the analysis results are not affected by invalid data.

The cleaned data then often needs to be transformed or further processed to fit the format required for analysis, which can include normalization, aggregation, or data encoding. In the modeling stage, various algorithms and analysis techniques are applied to identify patterns and relationships in the data, with some common techniques used in Data Mining, such as classification, regression, Clustering, and association. Once the model is developed, the results of the analysis are evaluated to determine its accuracy and relevance, and these results are then interpreted to provide insights that can be used in decision making.

The results of Data Mining can be applied in various fields, such as marketing, finance, healthcare, and security, to improve business strategies, identify risks, and increase operational efficiency. By understanding patterns and trends in data, organizations can make better and more strategic decisions, improve efficiency, personalize services, and detect fraud, making Data Mining a very valuable tool in today's information age, where the volume of data generated continues to increase and can provide valuable insights from the data they have, which in turn can improve their performance and competitiveness in the market.

2.4. Clustering

Clustering is one of the techniques in Data Mining that aims to group data into groups based on certain characteristics in common. This technique is applied in various fields, such as customer segmentation, genetic data analysis, document grouping, and image and video processing. By using the Clustering process, complex data can be simplified into more structured groups, making it easier to understand the relationships between data and support more effective decision making. Each group produced by the Clustering algorithm reflects certain patterns or characteristics in the data, which can help in identifying trends, predicting behavior, or grouping similar entities for further analysis[17].

2.5. K-Means Clustering

The K-Means algorithm is one of the techniques in Data Mining that is used to perform Clustering or grouping data with an unsupervised approach, where the modeling process does not require predetermined labels or categories[18].

2.6. Random Forest Algorithm

The Random Forest algorithm is a further development of the Classification and Regression Tree (CART) algorithm. This algorithm was introduced by Leo Breiman and utilizes randomization techniques in the decision tree building process. In Random Forest, a number of decision trees are built using random subsets of the training data and available features. Each decision tree produces outputs that are then aggregated to produce the final predictions[19].The Random Forest algorithm also implements an approach known as bagging (Bootstrap Aggregating), where different features from the data set are assigned to each decision tree built. This feature randomization process, often referred to as "bootstrapping," allows each tree to be trained on a random subset of the data, thereby increasing the diversity of the resulting model. In this way, Random Forest can reduce the overfitting problem that often occurs in single decision tree models, because the random selection of features helps reduce the interdependence between features[20].

3. RESEARCH METHOD

This study uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as a framework in managing and implementing the data analysis process. This method consists of several systematic stages, starting from understanding the business to implementing the results, to ensure accurate and applicable research results. Figure 1 below shows the complete flow and stages in CRISP-DM which is the basis for implementing this research.

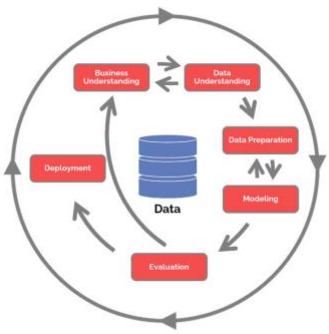

Fig. 1: CRISP-DM Stages

This study uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) approach consisting of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and implementation. The data used is secondary data from the Jayapura Police in 2020–2024 totaling 1172 cases of violence. The variables analyzed include the type of violence, number of reports (CT), case resolution status (SP3, ADR, P21), and number of completed cases (CC). After data cleaning and transformation, the Random Forest algorithm was used for classification, and K-Means Clustering for grouping. The model was evaluated using accuracy, precision, recall, and F1-score metrics for classification, and Silhouette Score and Elbow Method for clustering. The results of the model are visualized to support strategic decision making by law enforcement officers.

4. Results and Discussion

4.1. Data Understanding

This stage aims to understand the structure, quality, and characteristics of the data used in the study. The data used is data on cases of violence in the jurisdiction of the Jayapura Police during the period 2020 to 2024, with a total of 1,172 cases, 50 data entries. Each entry contains a number of important variables such as type of violence, number of case reports (CT), resolution status (ADR, SP3, P21, overflow), and the number of cases completed (CC) and the number still in process.

4.2. Data Preparation

This stage involves the process of cleaning and transforming data so that it is ready for use in analysis with the Random Forest and K-Means Clustering algorithms. This process is very important to ensure that the data used can provide accurate and reliable results in both algorithms.

4.2.1. Data Collection

At this stage, relevant data is collected from a trusted source, namely the Jayapura Police. Data collected as much as 50 data, namely violence data from 2020-2024. An example of the data that will be shown in **Table 1** below.

Table 1: Data on cases of violence in Jayapura Regency

Type of violence	Year	ct	adr	p21	overflow	sp3	cc	IN PROCESS
ORDINARY ASSAULT	2020	54	26	15	0	0	41	13
SERIOUS ASSAULT	2020	3	0	3	0	0	3	0
THREATENING / EXTORTION	2020	5	1	3	0	0	4	1
THEFT	2020	8	0	5	0	0	5	3
GANGING UP	2020	34	15	8	0	0	23	11
DESTRUCTION	2020	9	1	1	0	0	2	7
RAPE	2020	1	0	2	0	0	2	-1
DOMESTIC VIOLENCE	2020	18	5	2	0	0	7	11
MURDER	2020	1	0	0	0	0	0	1
THEFT	2020	1	0	0	0	0	0	1

These studies show the strong potential of the Random Forest and K-Means algorithms in data classification and clustering, but none have specifically integrated them for the analysis of cases of violence in certain areas such as Jayapura which is the focus of this study.

4.2.2. Data Cleaning

The data cleaning stage involves several steps, starting with the removal of records that lack complete information in the attributes: *jenis kekeran* (type of crime), year, CT, ADR, P21, *limpah* (transferred), SP3, CC, ongoing process, and completed/withdrawn cases. Incomplete data is considered less relevant as it may affect the accuracy of classification and clustering results. The researcher identified missing values in the dataset, as presented in **Table 2** below.

Table 2: Missing Values per Column

Column Name	Missing Values
Type of Violence	0
Year	0
CT	0
ADR	0
P21	0
Transferred (Limpah)	0
SP3	0
CC	0
Still in Process	0
Case Completed/Withdrawn	0
District	0
Sentani District	0
West Sentani District	0
East Sentani District	0

Initial exploration of the dataset revealed that there are no missing values, as all previously missing entries were filled with zeros during the data cleaning stage. Next, an analysis was conducted on the value distribution of each numerical feature. For example, the number of reported cases (CT) ranges from 1 to several dozen per type of violence. Additionally, an outlier check was performed. The results of outlier detection are presented in **Table 3**.

Table 3: Outlier Detection Results

Column	Outliers
Type of Violence	0
Year	0
CT	0

ADR	0
P21	0
Transferred (Limpah)	0
SP3	0
CC	0
Still in Process	0

4.2.3. Data Transformation

Data transformation is a critical stage in preprocessing, aiming to convert raw data into a more suitable format for further analysis. In this study, the transformation process was carried out through several main steps.
One of the key steps was data encoding, which transforms categorical attributes into numerical formats to be processed by machine learning algorithms. The encoding used in this research is presented in **Table 4.**

Table 4: Label Encoding Results

CT	ADR	P21	Transferred	SP3	CC	Still in Process
0	1.849691	1.366342	2.366432	0	0	2.108992
1	-0.852764	-0.780304	-0.169031	0	0	-0.717906
2	-0.746786	-0.697741	-0.169031	0	0	-0.639923
3	-0.587818	-0.780304	0.338062	0	0	-0.561939
4	0.789905	0.458145	1.098701	0	0	0.841762

The next step was feature importance analysis, which aims to identify the most influential attributes for classification or clustering out comes. This stage is essential for improving both the efficiency and accuracy of the Random Forest and K-Means algorithms applied in the study. The feature importance analysis is illustrated in **Figure 2.**

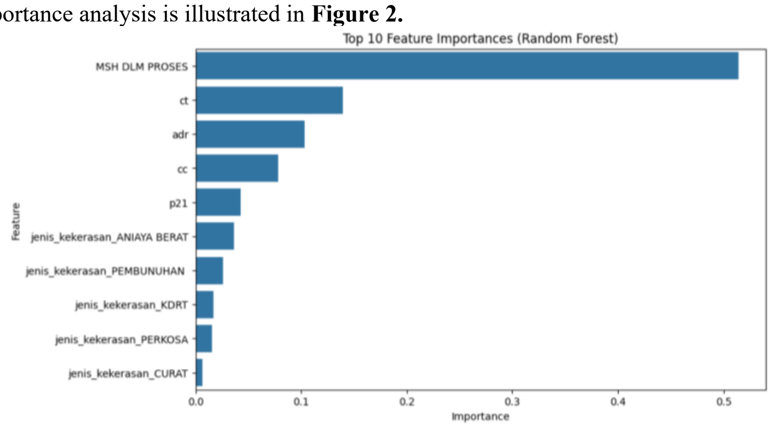

Fig. 2: Feature Important

4.2.4 Dimensionality Reduction

After completing the preprocessing stage, the data was analyzed using Principal Component Analysis (PCA) to reduce dimensionality without sacrificing significant information. In this study, two principal components were selected (**n_components = 2**), which together explain approximately **80.41%** of the total variance in the dataset. Selecting two components enables data representation in two dimensions, which facilitates visualization and supports more efficient clustering. The PCA transformation results are shown in **Figure 3**.

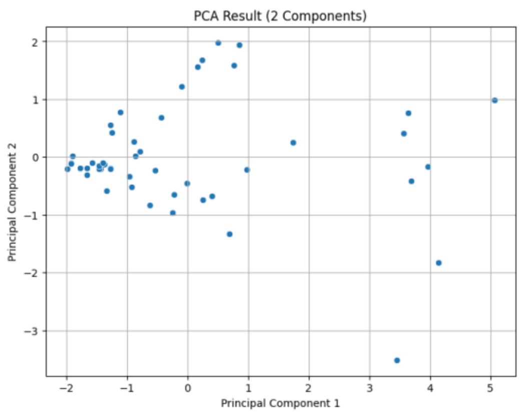

Fig. 3: Visualization of PCA Results Using Two Principal Components

4.2.5. Train-Test Data Splitting

Upon completing the preprocessing steps, the dataset was split into training and testing sets for model training and evaluation purposes. The variable df_processed was used as the feature set (X), while the column "MSH DLM PROSES > 0" was used as the target variable (y). The dataset was split using the train_test_split function, with **80% as training data** and **20% as testing data**. To ensure reproducibility and consistency of the results, random_state=42 was applied. Details of the dataset split are presented in **Table 3**.

Table 5: Dataset Split into Training and Testing Sets

Sub-set	Number of Rows	Number of Feature Columns	Description
Original Data	51	19	Before splitting
Training – Features	40	19	80% of the data
Testing – Features	11	19	20% of the data
Training – Target	40	1	Label “MSH DLM PROSES > 0”
Testing – Target	11	1	Label “MSH DLM PROSES > 0”

4.2. Modeling

This section presents the development of predictive models using two algorithmic approaches: **Random Forest** for classification and **K-Means** for clustering. These methods were selected due to their strong performance with diverse data types and alignment with the study's objectives.

4.2.1. Random Forest (Classification)

Random Forest is an ensemble-based classification algorithm that operates by building multiple decision trees and aggregating their predictions to improve accuracy and reduce overfitting. In this study, the algorithm was used to classify report statuses into two categories: **In Process** and **Not in Process**. Before training the model, the data was split using an 80:20 ratio. The model was trained using default parameters, with n_estimators = 100 and random_state = 42. The evaluation results indicated strong performance, achieving an accuracy of **91%**, as confirmed by the confusion matrix. The model successfully classified most instances correctly, including **all In Process** cases.

4.2.2. K-Means (Clustering)

K-Means is a clustering algorithm that aims to group data into clusters based on feature similarity. Prior to its application, dimensionality reduction via PCA was performed to ease visualization and enhance computational efficiency. The optimal number of clusters was determined using the **Elbow Method**. From the resulting plot, the elbow point was identified at **K=3**, hence three clusters were used. The data was then grouped into three distinct clusters, as illustrated in **Figure 4**.

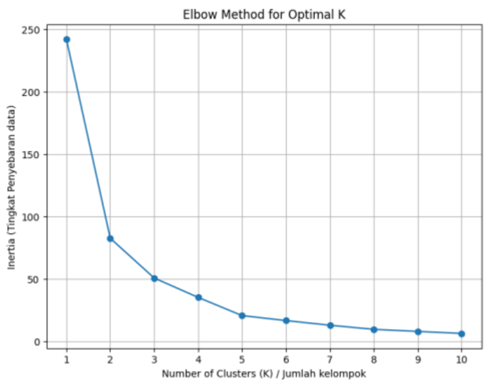

Fig. 4: Result of the Elbow Method

4.3. Evaluation

The final stage involved model evaluation and interpretation, which are critical to assess how well the analysis outcomes align with the research objectives. The Random Forest classifier was evaluated using metrics such as **accuracy**, **precision**, **recall**, and **F1-score**. The model achieved strong results, with **accuracy of 0.91**, as shown in **Table 4**.

Table 6: Evaluation Results of Random Forest Model

Label	Precision	Recall	F1-Score	Support
Not Processed (0)	1.00	0.75	0.86	4

In Process (1)	0.88	1.00	0.93	7
Accuracy			**0.91**	**11**

The **confusion matrix**, as shown in **Figure 5**, illustrates the classification accuracy across target categories.

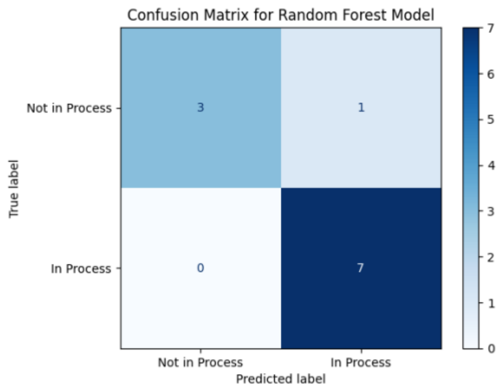

Fig. 5: Visualization of the Confusion Matrix

K-Means clustering was evaluated using the **Silhouette Score**, which measures how similar an object is to its own cluster compared to others. The score ranges from -1 to 1, with values closer to 1 indicating better clustering. In this study, the highest Silhouette Score was **0.56** at **K=3**, suggesting reasonably well-separated clusters. The results are shown in **Table 5**.

Table 7: Silhouette Score Results

Number of Clusters	Silhouette Score
2	0.51
3	0.56
4	0.53

4.3.1 Time Series Forecasting

Violence incident data was further analyzed using **Holt-Winters Exponential Smoothing** to forecast future trends. Historical incident data was processed to calculate monthly incident counts. This is illustrated in **Figure 6**.

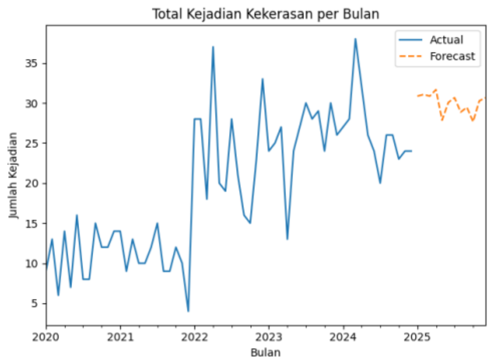

Fig. 6: Number of Incidents per Month

The Holt-Winters model was applied to forecast the total number of violence incidents per month for the next 12 periods (2025). Forecast results are detailed in **Table 6**.

Table 8: Forecasted Total Violence Incidents per Month (2025)

Date	Forecast_Total
01/01/2025	3.085.609
02/01/2025	3.105.598
03/01/2025	3.085.592
04/01/2025	3.165.597
05/01/2025	2.785.606
06/01/2025	3.005.566
07/01/2025	3.065.626
08/01/2025	2.885.580

09/01/2025	2.945.581
10/01/2025	2.765.640
11/01/2025	3.025.568
12/01/2025	3.065.604

The forecast is presented in a time series graph comparing actual data to projected results, enabling month-by-month trend analysis. Additionally, forecasting was performed for the three most frequently occurring types of violence. Their trends are illustrated in **Figure 7**.

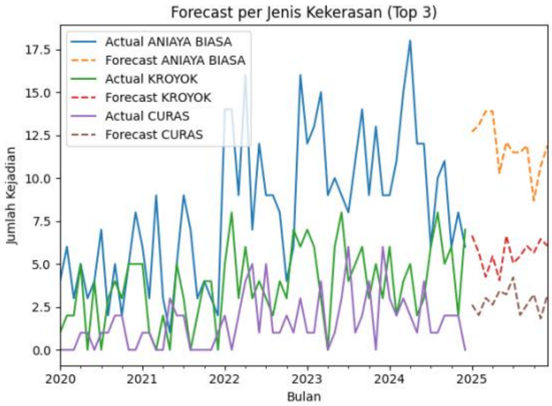

Fig. 7: Forecast Analysis of the Three Most Frequent Types of Violence

Forecasting was also conducted based on **incident locations**, allowing the identification of location-specific patterns. This analysis is shown in **Figure 8**.

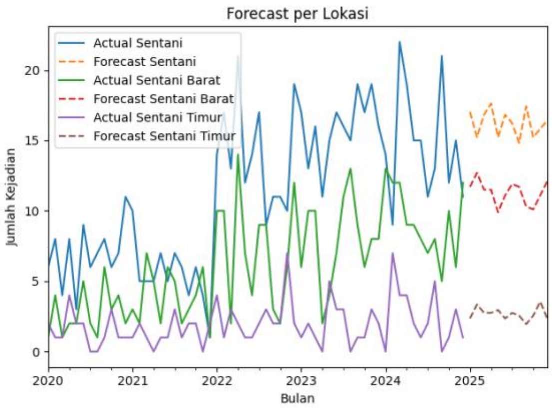

Fig. 8: Location-Based Forecasting

4.4. Deployment

This section outlines how the analytical findings can be deployed to support case handling and policy-making. The observed trend of increasing violence cases from 2020 to 2024—particularly the surge between 2021 and 2022—can be reported routinely to assist resource allocation, preventive policy formulation, and performance evaluation.

The clustering results, which identified **"ANIAYA BIASA"** and **"CURAS"** as dominant types in priority clusters, can guide automated case prioritization, dashboard development, and the creation of handling protocols. Additionally, key features from the Random Forest model—such as **CT**, **ADR**, **P21**, and **CC**—can be integrated into standard operating procedures (SOPs) via evaluation checklists, officer training, and early warning systems.

These findings should be effectively communicated to stakeholders through executive summaries, presentations, or publications to ensure actionable use and strengthen the violence case management system.

5. Conclusion

Based on a series of data analyses on violence cases from 2020 to 2024—including data cleaning, outlier handling, transformation, dimensionality reduction, classification and clustering modeling, as well as trend and forecasting analysis—several key conclusions can be drawn.

First, the trend analysis reveals a consistent increase in reported violence cases over the years, with a significant spike between 2021 and 2022. Forecasting using the Holt-Winters Exponential Smoothing method predicts that this upward trend will likely continue into 2025. The model identifies recurring seasonal patterns and a persistent growth trajectory, indicating the need for heightened vigilance and proactive intervention in the coming years.

Second, through the application of K-Means clustering on dimensionally reduced data (via PCA), the cases were grouped into three distinct clusters with varying characteristics. Notably, Cluster 1—dominated by "common assault" cases—and Cluster 2—characterized by "robbery with violence"—show a significantly higher number of reports and a larger proportion of unresolved cases compared to Cluster 0. These findings provide a solid basis for prioritizing resources and actions toward the most pressing case types.

Third, the Random Forest model's feature importance analysis highlights several key attributes that strongly influence the likelihood of case resolution delays. Apart from the unresolved status variable itself (“MSH DLM PROSES”), the most impactful features include the number of reports (CT), the status of Alternative Dispute Resolution (ADR), the completeness of case files (P21), and the stage of prosecution (CC). Understanding and managing these factors can support faster case handling and improved judicial processes.

Finally, the practical implications of this study are significant. The growing trend of violence and the identified high-priority clusters require careful allocation of resources and targeted preventive strategies—especially in periods and regions with heightened risk. Moreover, insights into the predictive factors for prolonged case processing can inform the refinement of Standard Operating Procedures (SOPs), officer training, and early-warning systems. The integration of classification, clustering, and forecasting approaches has produced a comprehensive and actionable analysis framework.

Overall, this research not only presents a quantitative overview of violence trends and projections but also offers data-driven insights for more efficient and effective case management strategies. The recommendations derived from this study are expected to contribute meaningfully to reducing violence and improving the justice system.

Acknowledgment

The authors would like to express their sincere gratitude to all individuals and institutions who supported this research. Special thanks are extended to the academic advisors for their valuable guidance, constructive feedback, and continuous encouragement throughout the research process. The authors are also thankful to the Jayapura Police Department (Polres Jayapura) for providing the necessary data and insights that significantly contributed to the analysis. Appreciation is further extended to colleagues and family members whose moral support and motivation were instrumental during the completion of this study. Every form of assistance—be it in the form of time, resources, or intellectual input—is deeply acknowledged and appreciated. This research was not funded by any external grant.

References

[1] D. Winarti, M. Kom, E. Revita, and M. Kom, “Penerapan Data Mining untuk Analisa Tingkat Kriminalitas Dengan Algoritma Association Rule Metode FP-Growth,” *J. SIMTIKA*, vol. 4, no. 3, pp. 8–22, 2021.
[2] M. Riskandi, M. Martanto, and U. Hayati, “Klasterisasi Korban Kekerasan Menggunakan Algoritma K-Means Di Jawa Barat,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 820–826, 2024, doi: 10.36040/jati.v8i1.8457.
[3] W. Dari, S. Aliyah, and R. I. Setiyawati, “Penerapan Aplikasi Microsoft Excel Dalam Pengelolaan Data Nilai Siswa Pada TK Kartika I-1 Medan Helvetia Application,” *J. Pengabdi. Masy. Sains dan Teknol.*, vol. 2, no. 2, pp. 198–205, 2023.
[4] Trivusi, “RANDOM FOREST,” *trivusi*, 2022, [Online]. Available: https://www.trivusi.web.id/2022/08/algoritma-random-forest.html
[5] F. Putri, “K-MEANS”, [Online]. Available: https://dibimbing.id/blog/detail/apa-itu-k-means-clustering-kelebihan-proses-contoh
[6] A. Chofidah and S. Pramana, “Mengungkap Lanskap Kejahatan Provinsi di Indonesia Tahun 2021 : Analisis Perbandingan K-Means dan Logika Fuzzy,” vol. 2024, no. Senada, pp. 421–433, 2024.
[7] M. A. K-means, W. Kurniawan, F. R. Pradhana, and K. A. Zen, “Analisis Clustering Kasus Bunuh Diri di Jawa Tengah dengan,” vol. 9, no. 2502, pp. 47–55, 2024.
[8] D. I. Wilayah and K. Depok, “Implementasi algoritma k-means untuk clustering kasus kriminal di wilayah kota depok 123,” vol. 2, no. 1, pp. 408–415, 2025.
[9] Fadil Danu Rahman, M. I. Z. Mulki, and A. Taryana, “Clustering Dan Klasifikasi Data Cuaca Cilacap Dengan Menggunakan Metode K-Means Dan Random Forest,” *J. SINTA Sist. Inf. dan Teknol. Komputasi*, vol. 1, no. 2, pp. 90–97, 2024, doi: 10.61124/sinta.v1i2.15.
[10] Nurdiyanto Yusuf, “Prediksi Produksi Daging Sapi Di Indonesia Menggunakan Random Forest Regression: Analisis Data 2018-2025,” *J. Ilm. Tek.*, vol. 3, no. 2, pp. 134–142, 2024, doi: 10.56127/juit.v3i2.1620.
[11] H. Sa’diah, U. Enri, and T. Nur Padilah, “Penerapan Algoritme K-Means Dalam Segmentasi Daerah Rawan Kekerasan Anak Di Jawa Barat,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1351–1357, 2023, doi: 10.36040/jati.v7i2.6838.
[12] K. Luana, P. Widodo, H. Faridah, and U. S. Karawang, “Jurnal Panorama Hukum,” pp. 126–138, 2021.
[13] U. Medan and M. bukan mampus/Semester 7/refrensi/jurnal/BAB I. pd. Area, “Tinjauan Umum Tentang Kekerasan,” *Penyakit Kanker*, no. 1, pp. 1–12, 2018.
[14] Y. B. Ferdiansyah, “Penerapan Metode K-Means Untuk Clustering Tingkat Kejahatan di Indonesia,” 2023.
[15] A. Latuheru, “Faktor-Faktor Yang Mempengaruhi Pertumbuhan Ekonomi Di Kota Jayapura,” vol. 16, no. 1, 2024.
[16] F. Salsabila, T. Ridwan, and H. H, “Analisa Volume Penyebaran Sampah Di Karawang Menggunakan Algoritma K-Means Clustering,” *J. Inform.*

dan Tek. Elektro Terap., vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4226.
[17] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26
[18] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, “Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
[19] N. Muna *et al.*, “Penerapan Algoritma Random Forest Untuk MEMPREDIKSI JUMLAH SANTRI BARU,” vol. 5, no. 1, pp. 49–54, 2024.
[20] H. Oktavianto, H. W. Sulistyo, G. Wijaya, D. Irawan, and G. Abdurrahman, “Analisis Komparasi Kinerja Metode Decision Tree dan Random Forest dalam Klasifikasi Teks Data Kesehatan,” *Bina Insa. ICT J.*, vol. 11, no. 1, pp. 56–65, 2024, [Online]. Available: https://www.kaggle.com/datasets/falgunipatel19/biomedical-text-publication-classification.