

# *Analysis of Public Perception on Domestic Violence Cases using Support Vector Machine Algorithm*

<sup>1</sup>Mirdatul Husnah\*, <sup>2</sup>Rahmat Hidayat

<sup>1</sup>Department of Information System, FMIPA, Universitas Riau

<sup>2</sup>Department of Informatics Management, FMIPA, Universitas Riau

<sup>1,2</sup>Kampus Bina Widya KM. 12,5, Simpang Baru Pekanbaru Riau, Indonesia

\*e-mail: [mirdatulhusnah@lecturer.unri.ac.id](mailto:mirdatulhusnah@lecturer.unri.ac.id)

(received: 29 October 2024, revised: 14 November 2024, accepted: 18 November 2024)

## *Abstract*

Domestic violence is currently a case that is easily exposed by the public. People can easily find many cases through social media. The latest case was experienced by a social media influencer, Cut Intan. This case has attracted public attention and is widely discussed on several social media, one of which is on the X app. With this phenomenon, an analysis of public sentiment towards domestic violence cases that occur in Indonesia is needed. The analysis was conducted using the Support Vector Machine algorithm, a classification algorithm that can classify values into certain classes and has a good level of accuracy. Experiments on analyzing public sentiment towards domestic violence cases using the SVM algorithm resulted in an accuracy score of 95%. The precision score for negative sentiment is 94%, neutral sentiment is 100%, and positive sentiment is 100%. The recall result for negative sentiment is 100%, neutral sentiment is 67%, and positive sentiment has a value of 77%. The results of the f-1 score on negative sentiment are 97%, 80% neutral sentiment, and 87% positive sentiment. While the percentage of community sentiment obtained is 84.40% having negative sentiment, 8.24% having positive sentiment, and 7.36% having neutral sentiment.

**Keywords:** KDRT, support vector machine, sentiment analysis.

## **1 Introduction**

Family is the smallest social unit in society and is the first place for humans to start life and establish relationships. A loving and harmonious family relationship is certainly the dream of every family. However, nowadays there are many complicated conflicts that occur due to incompatibility or misunderstanding between each other and family members [1]. The conflict that often occurs in families is none other than violence. The number of cases of violence in Indonesia recorded by the Ministry of Women's Empowerment and Child Protection from January 1, 2024, to August 16, 2024, has occurred in as many as 15,658 cases. The demographic data of violence, when viewed based on the place of occurrence, is divided into 9,600 cases occurring in households, 1,602 occurring in public facilities, 852 occurring in schools, 204 in the workplace, 16 cases in short-term educational institutions, and 3,367 cases in other places. Then, based on the gender of the perpetrators of violence, there were 10,417 male perpetrators and 1,315 female perpetrators [2]. So, based on these demographic data, men are the most common perpetrators of violence, and the household is the place where most violence occurs. Domestic violence is a common and serious social problem and relates to physical, psychological, or emotional abuse inflicted by someone in a domestic setting [3].

Domestic violence has become one of the issues in the social phenomenon of households that often occur throughout the world, including in Indonesia [4]. One of the main causes of domestic violence is the lack of knowledge and understanding of the rights and obligations of spouses [5]. Domestic violence often targets women, particularly due to the patriarchal stigma that views women as powerless. Thus, men are often the perpetrators of violence in domestic violence cases [6]. In Indonesia, Law No. 23 of 2004 has regulated the Eradication of Domestic Violence (PKDRT), which should provide protection to women who are victims of domestic violence [7]. Domestic Violence Article 1 Number 1 states that "Domestic violence is any act against a person, especially women, which results in physical, sexual, psychological misery or suffering, and/or neglect of the household, including threats to commit acts, coercion, or unlawful deprivation of liberty within the scope of the household." [8]. However, in reality, such protection is still difficult to implement, such as one recent

<http://sistemasi.ftik.unisi.ac.id>

incident that surfaced on social media and was experienced by celebrity Cut Nabila. This case provides an illustration of the struggle of a victim of domestic violence who seeks protection for herself and requires legal handling in following up on the perpetrator of domestic violence. This case has attracted a lot of public attention and has been widely discussed on social media, one of which is X.

X is a social media brand renewal from the previous name, Twitter. X is a microblogging and social networking platform that focuses on short and concise messages and tweeting ideas, namely sharing thoughts and short updates with other users [9]. In general, X is used to share funny things, quote tweets, and repost in order to interact with other users, such as liking, responding, or sharing further [10]. The large and diverse user base of X makes X a platform rich in data sources for obtaining and understanding the dynamics of public opinion that are attractive to the public [11]. Sentiment analysis is a branch of data mining that is used to analyze, process, and obtain textual data from entities such as services, products, individuals, organizations or specific topics [12]. In this study, the Support Vector Machine (SVM) algorithm will be used to analyze public sentiment regarding domestic violence cases. The SVM algorithm was chosen because this algorithm can classify values into certain classes and has a better level of accuracy than other algorithms [13]. With this SVM algorithm, it is expected to be able to identify dominant sentiments and understand how information trends regarding domestic violence cases develop in X.

Thus, this research on sentiment analysis aims to provide insight into public opinion on domestic violence cases and how these cases affect public opinion. This research also aims to evaluate the level of accuracy of the SVM algorithm in classifying sentiments on domestic violence cases from data X. The benefits of this research can contribute to understanding how information about domestic violence cases is spread on social media and how the public reacts to this information. Thus, this research is expected to be a reference in managing and responding to information about domestic violence cases circulating on social media.

## **2 Literature Review**

Previous research on sentiment analysis using the SVM algorithm has been done a lot before, such as one by Rian Tineges, Agung Triayudi, and Ira Diana Sholihati in 2020, entitled "Sentiment Analysis of Indihome Services Based on Twitter with the Support Vector Machine (SVM) Classification Method." The study explains the SVM algorithm in analyzing sentiment regarding the opinions of Indihome service users on Twitter. This study used 450 tweet data, with test results of 18.4% for positive sentiment and 81.6% for negative sentiment. Thus, the number of positive sentiments was 83 tweets, and negative sentiments were 367 tweets. Then, it was evaluated against the test data to determine the level of predicted data error of 13%. Thus, the level of accuracy given is 87% accuracy, 86% precision, 95% recall, and 90% f1-score [14].

Then, another study on sentiment analysis using SVM was conducted by Primandani Arsi and Retno Waluyo in 2021, entitled "Sentiment Analysis of Discourse on Moving the Indonesian Capital City Using the Support Vector Machine (SVM) Algorithm." The results of the study provide sentiment classification results into two classes, namely positive and negative. The results of the test carried out on the sentiment of moving the capital city as many as 1236 tweets with a division of 404 positive and 832 negative gave an accuracy result of 96.68%, recall of 94.04%, and AUC of 0.979% [15].

Then, another study on sentiment analysis and discussing domestic violence and sexual harassment was conducted by Brescia Ayundina Yuniarossy, Kartika Maulida Hindrayani, and Aviolla Terza Damaliana in 2024, entitled "Sentiment Analysis of Feminism Issues on Twitter Using the Convolutional Neural Network (CNN) Model." This study aims to identify sentiment patterns towards Twitter users regarding feminism issues, especially domestic violence and sexual harassment. The results of this study indicate that cases of sexual harassment receive higher sentiment than cases of domestic violence. With an accuracy value of 86% for sexual harassment and 82% for domestic violence [16].

Thus, based on previous research, it is known that sentiment analysis in the case of Indihome services and the relocation of the capital using the SVM algorithm often provides higher accuracy results, compared to domestic violence cases using CNN. And research on sentiment analysis of

domestic violence cases using the SVM algorithm has not been carried out by previous researchers. Especially the domestic violence case discussed in this study concerns the case of one of the celebrities that shocked the public, so the sentiment in this study will focus on the domestic violence case of the celebrity. Therefore, researchers are interested in raising the topic of public sentiment analysis regarding domestic violence cases using the SVM algorithm.

### 3 Research Method

Methodical approach was developed in order to fully meet the objectives of this study. This methodology ensures each step, from data collection to the final analysis, is meticulously designed for clarity and accuracy. Figure 1 is an illustration of the complete research method framework employed in this study.

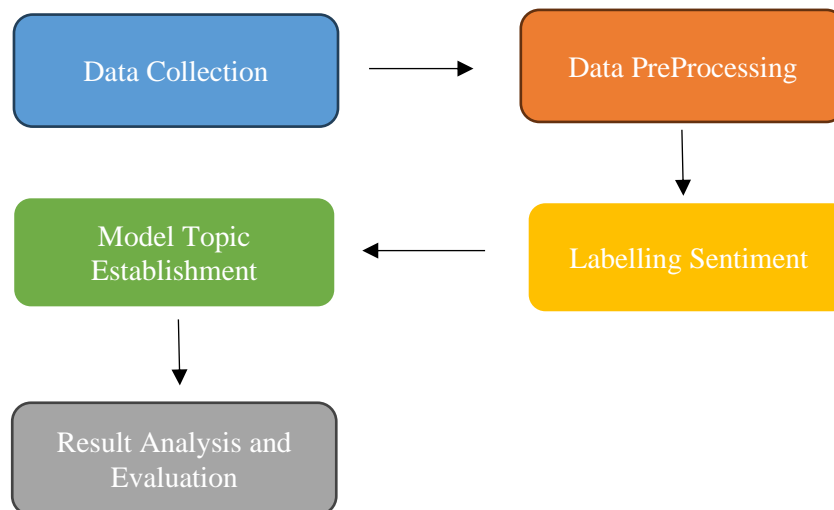


Figure 1. Research method

Figure 1 Research Method displays the sequence of phases involved, illustrating how each step interconnects to support the progression of the research process. The initial stage is to collect data about domestic violence cases through social media X. This data collection stage is carried out by crawling using Python programming language. The crawled data are tweets that have the keyword domestic violence from August 15, 2024, to August 18, 2024, with a total of 1252 data points, where discussions about the domestic violence case of celebrity Cut Intan are currently being widely discussed on social media X to become a trending topic on social media X. At this stage, the data that has been obtained is stored in CSV format, which will then be used in the data labeling stage.

Before labeling, it will enter the pre-processing stage. The preprocessing stage consists of various processes such as case folding, filtering, tokenizing, stopword removal, and stemming. Case folding is the stage of changing all data to lowercase, then filtering is the stage of cleaning data from punctuation or symbols or URLs that are not needed, then the tokenizing stage is the stage of breaking text into words, the next stage is the stopword removal stage is the stage of removing words that often appear but do not buy important information such as "yang", "this", "that", "and", etc., and the stemming stage is the stage that changes words to their basic form. Then, it will enter the data labeling stage with data that has been cleaned in the preprocessing stage, then the data will be divided based on negative, positive, and neutral sentiments. The sentiment labeling will be based on the score obtained for each data. If the score  $\geq 0.05$  then it is included in positive sentiment, if the score obtained = 0 then the sentiment becomes neutral, and other than the two provisions it will become negative sentiment.

The next stage will be classification using the support vector machine (SVM) algorithm. Python will train the model using training data, and then the results will be predicted using test data. After testing, the model evaluation will be carried out by providing accuracy results and reports of precision, recall, f1-score, and support. After obtaining the accuracy of the results from SVM, then

visualization will be carried out using a diagram that will show the percentage of each sentiment, confusion matrix, and wordcloud.

#### 4 Results and Analysis

The data obtained from the crawling process were 1252 tweets based on data collection since August 15, 2024. The reason for choosing August 15, 2024, was because since that date, domestic violence has become a trending topic on the X application and has become the number 1 most tweeted due to the domestic violence case experienced by celebrity Cut Intan. The results of the data crawling are shown in Figure 2.

	conversation_id_str	created_at	favorite_count	full_text	id_str
0	1823509705156911247	Tue Aug 13 23:59:27 +0000 2024	0	Selingkuh dan kdrt itu engga bgtttttt knp ada ...	1823509705156911247
1	1823506193664630915	Tue Aug 13 23:59:15 +0000 2024	0	@aidenxvx takut kdrt dan selingkuh	1823509655789637977
2	1823509602920751616	Tue Aug 13 23:59:03 +0000 2024	0	Sulit maksa orang yg kdrt buat berobat. Apalag...	1823509602920751616

Figure 2. Crawling data

After carrying out the crawling process, it will then enter the first preprocessing stage, namely case folding, which will change all data into lowercase letters in Table 1.

Table 1. Case folding

Before Case Folding	After Case Folding
Sulit maksa orang yg kdrt buat berobat. Apalagi kalo pelaku kdrt jg ternyata dia selalu merasa dirinya benar. Apesnya malah ternyata dia NPD dan baru ketauan karena korban memberanikan diri ke psikolog dan psikiater.	sulit maksa orang yg kdrt buat berobat. apalagi kalo pelaku kdrt jg ternyata dia selalu merasa dirinya benar. apesnya malah ternyata dia npd dan baru ketauan karena korban memberanikan diri ke psikolog dan psikiater.

After the case folding stage, then filtering will be carried out, which is the stage of cleaning data from unnecessary punctuation, symbols, or URL. The filtering results can be seen in Table 2.

Table 2. Filtering

Before Filtering	After Filtering
sulit maksa orang yg kdrt buat berobat. apalagi kalo pelaku kdrt jg ternyata dia selalu merasa dirinya benar. apesnya malah ternyata dia npd dan baru ketauan karena korban memberanikan diri ke psikolog dan psikiater.	sulit maksa orang yg kdrt buat berobat apalagi kalo pelaku kdrt jg ternyata dia selalu merasa dirinya benar apesnya malah ternyata dia npd dan baru ketauan karena korban memberanikan diri ke psikolog dan psikiater

The next stage, after the filtering stage, is tokenizing, which is the stage to break the text into words. The tokenizing results can be seen in Table 3.

Table 3. Tokenizing

Before Tokenizing	After Tokenizing
sulit maksa orang yg kdrt buat berobat	["sulit", "maksa", "orang", "yg", "kdrt",

apalagi kalo pelaku kdrt jg ternyata dia selalu merasa dirinya benar apesnya malah ternyata dia npd dan baru ketauan karena korban memberanikan diri ke psikolog dan psikiater	"buat", "berobat", "apalagi", "kalo", "pelaku", "kdrt", "jg", "ternyata", "dia", "selalu", "merasa", "dirinya", "benar", "apesnya", "malah", "ternyata", "dia", "npd", "dan", "baru", "ketauan", "karena", "korban", "memberanikan", "diri", "ke", "psikolog", "dan", "psikiater"]
--	--

After tokenizing, the stopword removal stage will be carried out, which is the stage of removing words that often appear but do not contain important information. The results of stopword removal can be seen in Table 4.

**Table 4. Stopword removal**

Before Stopword Removal	After Stopword Removal
["sulit", "maksa", "orang", "yg", "kdrt", "buat", "berobat", "apalagi", "kalo", "pelaku", "kdrt", "jg", "ternyata", "dia", "selalu", "merasa", "dirinya", "benar", "apesnya", "malah", "ternyata", "dia", "npd", "dan", "baru", "ketauan", "karena", "korban", "memberanikan", "diri", "ke", "psikolog", "dan", "psikiater"]	["sulit", "maksa", "orang", "kdrt", "buat", "berobat", "apalagi", "kalo", "pelaku", "kdrt", "ternyata", "selalu", "merasa", "dirinya", "benar", "apesnya", "malah", "ternyata", "npd", "baru", "ketauan", "karena", "korban", "memberanikan", "diri", "psikolog", "psikiater"]

Then, after doing stopword removal, the stemming stage will be carried out, which is a stage that will change words to their basic form. The stemming results can be seen in Table 5.

**Table 5. Stemming**

Before Stemming	After Stemming
["sulit", "maksa", "orang", "kdrt", "buat", "berobat", "apalagi", "kalo", "pelaku", "kdrt", "ternyata", "selalu", "merasa", "dirinya", "benar", "apesnya", "malah", "ternyata", "npd", "baru", "ketauan", "karena", "korban", "memberanikan", "diri", "psikolog", "psikiater"]	["sulit", "maksa", "orang", "kdrt", "buat", "obat", "apalagi", "kalo", "pelaku", "kdrt", "nyata", "selalu", "rasa", "diri", "benar", "apes", "malah", "nyata", "npd", "baru", "tau", "karena", "korban", "berani", "diri", "psikolog", "psikiater"]

After the preprocessing stage is done, the data will be labeled. The preprocessed data will be divided based on positive, negative, or neutral sentiment. The labeling results can be seen in Table 6.

**Table 6. Labelling**

Text	Labeling
sulit maksa orang kdrt buat obat apalagi kalo pelaku kdrt nyata selalu merasa diri benar apes malah nyata dia npd baru tau karena korban berani diri psikolog psikiater	Negatif
kenapa allah suruh sabar syukur dalam hidup karena nyata sabar syukur berat ketika kasih ujian tidak enak harus sabar	Positif
salam malaysia nak tanya KDRT apa?	Netral

The next stage, after obtaining sentiment on each data, is that the classification will be carried out using the SVM algorithm. Before being tested using SVM, the data will be divided into training data

and testing data, with a total of 20% training and 80% testing. After classification with SVM, the accuracy results are obtained as in Figure 3. The accuracy result obtained is 0.952. For negative sentiment, the precision value is 0.94, recall is 1.00, and f-1 score is 0.97. Then for neutral sentiment, the precision value is 1.00, recall is 0.67, and f-1 score is 0.80. For positive sentiment, the precision value is 1.00, then recall is 0.77, and f-1 score is 0.87. Then, after obtaining the accuracy value from the labeling results, we will visualize the sentiment labeling results using a pie chart. The percentage of each data point will be grouped based on negative, positive, and neutral sentiment. A visualization of each sentiment will show the magnitude of the division between sentiments.

Akurasi: 0.952

Laporan Klasifikasi:

	precision	recall	f1-score	support
negatif	0.94	1.00	0.97	103
netral	1.00	0.67	0.80	9
positif	1.00	0.77	0.87	13
accuracy			0.95	125
macro avg	0.98	0.81	0.88	125
weighted avg	0.95	0.95	0.95	125

Figure 3. SVM algorithm accuracy

Based on Figure 4, it can be seen that negative sentiments obtained the largest percentage result in domestic violence sentiment of 84.40%, followed by positive sentiments with a total of 8.24%, and neutral sentiments obtained the smallest result of 7.36%. The results of this visualization show that X users gave more negative sentiments in domestic violence cases that were the topic of conversation at that time, followed by positive sentiments. Positive sentiments tend to contain prayers and good wishes for victims of domestic violence. Then the last is that neutral sentiments in domestic violence cases are very few. This neutral sentiment shows that the words of the tweet do not lead to negative or positive sentiments. Thus, this shows that X users are more likely to not support and tend to give negative sentiments towards domestic violence cases.

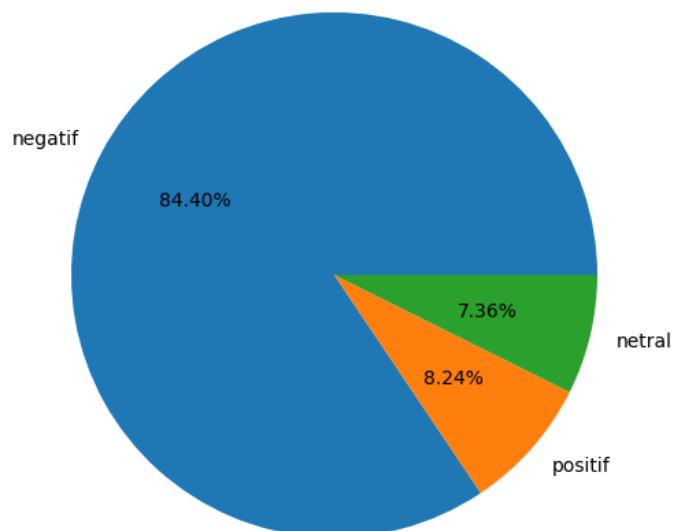


Figure 4. Pie chart sentiment

Furthermore, the sentiment results will also be visualized using a confusion matrix. In the confusion matrix-SVM, a comparison between the predicted labeling and the actual label can be seen. The results of the confusion matrix can be seen in Figure 5.



The visualization results using the confusion matrix in Figure 5 show that the actual negative and predicted negative values are 103, the actual negative and predicted neutral values are none, and the actual negative and predicted positive values are none. Then, the actual neutral and predicted negative values amount to 3, the actual neutral and predicted neutral values amount to 6, and the actual neutral and predicted positive values do not exist. Furthermore, the true value of positive and predicted negative is 3, the true value of positive and predicted neutral is 0, and the true value of positive and predicted positive is 10.

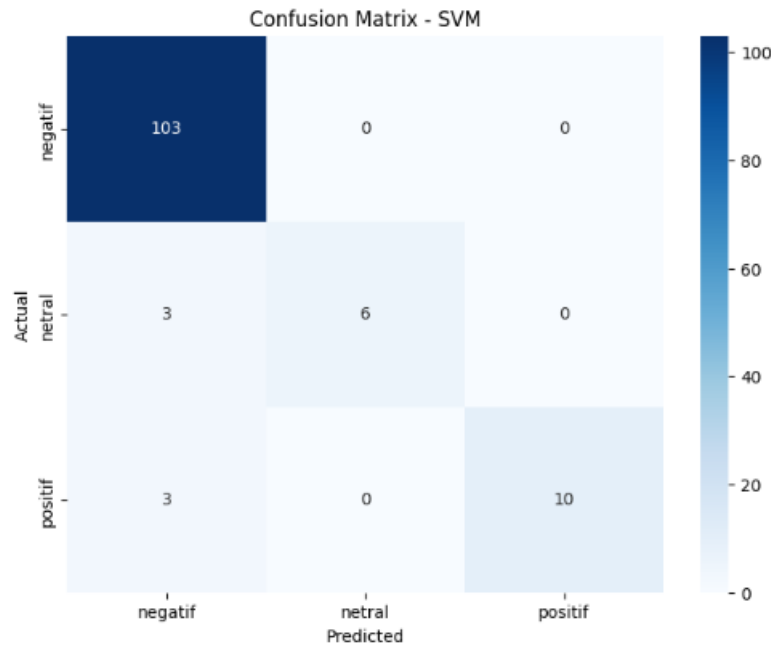


Figure 5. Confusion matrix SVM

The next visualization can be seen through the word cloud in Figure 6, which can show the most talked-about words in domestic violence cases. The more people talk about it, the bigger the word will be in the word cloud, and vice versa. As can be seen in Figure 6, words such as Domestic Violence, Children, and Victims of Domestic Violence are the most mentioned words in domestic violence cases, as well as Cut Intan Nabila, who was a victim in the domestic violence case in this study.



Figure 6. Word cloud visualization

Thus, this study shows that in cases of domestic violence, X users provide more negative sentiments, such as angry sentences directed at the perpetrators of domestic violence, then followed by neutral sentiments which usually refer to sentences that do not show positive or negative

sentiments related to cases of domestic violence, while positive sentiments usually refer to good sentences and prayers directed at victims of domestic violence.

## 5 Conclusion

This study conducted a sentiment analysis of the community towards domestic violence cases on the social media application X. Based on 1252 tweets from the community on the X application, it was found that 84.40% had negative sentiment, 8.24% had positive sentiment, and 7.36% had neutral sentiment. The classification results using the SVM algorithm produced an accuracy score of 95%. The precision score for negative sentiment was 94%, neutral sentiment was 100%, and positive sentiment was 100%. The recall result for negative sentiment was 100%, neutral sentiment was 67%, and positive sentiment was 77%. The f-1 score result for negative sentiment was 97%, 80% neutral sentiment, and 87% positive sentiment.

## Reference

- [1] R. Alimi and N. Nurwati, "Faktor Penyebab terjadinya Kekerasan dalam Rumah Tangga terhadap Perempuan." *Jurnal Pengabdian dan Penelitian Kepada Masyarakat (JPPM)*, pp. 20–27, 2024.
- [2] K. P. P. dan P. Anak, "Data Sistem Informasi Online Perlindungan Perempuan dan Anak (SIMFONI PPA) Kemen PPPA," 2024, [Online]. Available: <https://kekerasan.kemenpppa.go.id/ringkasan>
- [3] S. P. Imanda, A. Putri, and Y. W. Pertiwi, "Psychological Well Being pada Korban Kekerasan dalam Rumah Tangga," *J. Publ. Ilmu Psikol.*, vol. 2, no. 3, pp. 261–274, 2024, [Online]. Available: <http://dx.doi.org/10.1016/j.encep.2012.03.001>
- [4] M Noor Farchan and Dian Alan, "Analisis Yuridis Tindak Pidana Kekerasan dalam Rumah Tangga yang Mengakibatkan Kematian," *J. Ris. Ilmu Huk.*, pp. 111–116, 2023, doi: 10.29313/jrih.v3i2.2998.
- [5] B. B. Leksono and A. Wibowo, "Kepolisian dan Penegakan Hukum Pidana terhadap Tindak Kekerasan dalam Rumah Tangga," vol. 22, no. 2, pp. 190–206, 2024.
- [6] G. A. T. Panjaitan and L. Nurhadiyanto, "Kajian Perempuan sebagai Pelaku KDRT ditinjau dari Perspektif Victim Precipitation," *IKRA-ITH Hum. J. Sos. dan Hum.*, vol. 7, no. 3, pp. 49–58, 2023, doi: 10.37817/ikraith-humaniora.v7i3.3326.
- [7] Jesica and R. Zubaedah, "Penghapusan Kekerasan dalam Rumah Tangga (KDRT) terhadap Perempuan dan Anak," *J. Ilm. Wahana Pendidik.*, vol. 10, no. 11, pp. 308–315, 2024.
- [8] L. N. R. Indonesia, "Undang-Undang Penghapusan Kekerasan dalam Rumah Tangga," *UU PKDRT*, no. 23, 2004.
- [9] K. Wahyuningsih, "Pemanfaatan Aplikasi Twitter dalam Menulis Teks Berita," *Pros. Senada (Seminar Nas. Daring)*, vol. 2, no. 1, pp. 998–1003, 2022.
- [10] M. Dynel, "The Pragmatics of Sharing Memes on Twitter.pdf." *Journal of Pragmatics*, Elsevier, pp. 100–115, 2024.
- [11] R. N. Pamungkas, D. Permadi, and I. D. Florina, "Strategi Humor Gibran Rakabuming dalam Komunikasi Politik di Media Sosial X (Twitter)," *Pemerintah. dan Polit.*, vol. 9, no. 3, pp. 175–182, 2024.
- [12] B. Liu, "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition," *Sentim. Anal. Min. Opin. Sentim. Emot. Second Ed.*, no. May, pp. 1–432, 2020, doi: 10.1017/9781108639286.
- [13] C. Chazar and B. Erawan, "Machine Learning Diagnosis Kanker Payudara menggunakan Algoritma Support Vector Machine," *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 12, no. 1, pp. 67–80, 2020, doi: 10.37424/informasi.v12i1.48.
- [14] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen terhadap Layanan Indihome berdasarkan Twitter dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [15] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [16] B. A. Yuniarossy, K. M. Hindrayani, and A. T. Damaliana, "Analisis Sentimen terhadap Isu



Feminisme di Twitter menggunakan Model *Convolutional Neural Network* (CNN),” *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 5, no. 1, pp. 477–491, 2024, doi: 10.46306/lb.v5i1.585.