E-ISSN : 2963-4946
Vol. 1 No. 12 September 2023
Asian Journal of Social and Humanities
https://ajosh.org/

Analysis of School Community Sentiment towards Personal Data Protection Law Using Support Vector Machine (SVM) Method

Gusti Fachman Pramudi[1], Gerry Firmansyah[2], Budi Tjahyono[3], Agung Mulyo Widodo[4]
[1,2,3,4] Universitas Esa Unggul, Indonesia
Email: gusti.fachman@gmail.com

* Correspondence: gusti.fachman@gmail.com

KEYWORDS
School Community Sentiment, Personal Data Protection Law, Support Vector Machine (SVM)

ABSTRACT
This research aims to analyze the awareness status of the school community regarding the right to personal data protection, test and analyze the sentiments of the school community towards the implementation of the Personal Data Protection Law and as a means of outreach regarding personal data protection laws in the world of education, especially school community besides that to find out whether the Support Vector Machine method can be used as a method in conducting Sentiment Analysis research. The results of this study can be concluded that only 38% of the school community knows about this regulation while the other 62% still don't know much about this rule. Then for the use of the Support Vector Machine method which has been carried out five (5) trials using different variations of training data and test data produces an average accuracy rate of 85.97% with the highest results on training data and test data 50% - 50% that is equal to 88.00% and the lowest result is in the experiment of training data and test data of 90% - 10% which is equal to 84.44%. For the school community's sentiment towards the Personal Data Protection Act, it was 56% or as many as 496 of 887 words. which shows a neutral response and 8% or as many as 72 out of 887 sentiment words show a negative response.

Introduction

Concerns about misuse of personal data can also be seen that a percentage of as many as 59% of internet users feel worried if their personal data is misused by companies or certain parties with purely profit motives which can harm the data owner. The increase in internet users is inseparable from public awareness of technology which demands convenience in the globalization era as a supporting factor for other activities, including the emergence of new forms of crimes.

Personal data in the 21st century era is almost a primary need, because the transition in the real world which is increasingly shifting to new things in the form of all-

visual forms makes it easier for all activities to be carried out. There is a term "crime is a product of society itself" which applies to the rapid development of information technology which creates new things in the world of law. The crime of using technology as an internet-based medium is emerging and growing in a society that makes it commonplace.

Regulations governing the protection of personal data in Indonesia are explicitly regulated in several laws, such as Law Number 10 of 1998 concerning Banking, Law Number 36 of 2009 concerning Health, Law Number 24 of 2013 concerning Population Administration, and Law Number 19 of 2016 concerning Amendments to Law Number 11 of 2008 concerning Information and Electronic Transactions. Article 26 Paragraph (1) of the ITE Law regarding information through electronic media containing personal data does not explain in detail and comprehensively the principles of personal data protection, rights and obligations for data owners and stakeholders or the government in processing and using personal data. . The elucidation of the law in this article only provides a general definition of personal rights. In Paragraph (2) you can see the consequences if there is a violation related to personal data which is only compensation in nature, the potential for a weak position of the owner of personal data can be seen when an action occurs that harms the owner of personal data, even the owner of personal data does not realize that he has been harmed and in the case of The role of the state is only passive.

In 2019, the Indonesian Internet Service Users Association (APJII) noted that 196.71 million people in Indonesia had accessed the internet. This figure covers 73.7% of the total 270 million population of Indonesia. Java is the island with the most users, reaching 55.7% of the population of internet users in Indonesia. The use of technology makes it easier for people to do various activities, from communication, transportation to digital transactions. The use of internet technology has implications for the vulnerability of users' personal data. Each user should be able to determine whether their data can be used and disseminated by social media managers or applications. the user also has the right to determine the requirements that apply in one community regarding the use of personal data.

Personal data regarding full names, e-mails, social media accounts and even account numbers are required by various application services, one of which is to ensure user legitimacy and service accuracy. However, there is no guarantee that the personal data will be protected from misuse. Contact numbers, bank accounts and home addresses can be used by parties with malicious intent, such as committing cell phone fraud, hacking into bank accounts and robbing homes.

Protection of privacy and personal data is a factor that determines the level of online trust. The lack of protection causes privacy data to be spread to irresponsible parties, which can be financially detrimental, even threatening the safety of the owner.

The research on the sentiment of the school community on the protection of personal data aims to find out the sentiment of the school community on the right to the protection of personal data. The collection of primary data is expected to reflect the school community's understanding of personal data, its misuse, awareness of the right to protect personal data (Ndruru, 2022).

The authors identify the problems in this study as follows: 1. Does the school community know about the Personal Data Protection Law that has been established by the government? 2. What is the sentiment of the school community regarding the Personal Data Protection Act? 3. How is the Support Vector Machine method applied in analyzing the sentiments of the school community regarding the Personal Data Protection Act?

Based on the background of the problems above, the formulation of the problem in this study is as follows: "How is the application of the Support Vector Machine method in determining the effectiveness of the application of the Personal Data Protection Act in the school community" and the objectives of this study are: 1. Test and analyze status of school community awareness of the right to personal data protection. 2. Testing and analyzing the sentiments of the school community towards the implementation of the Personal Data Protection Act 3. As a means of outreach regarding personal data protection laws in the world of education, especially the school community. 4. To find out whether the Support Vector Machine method can be used as a method in conducting Sentiment Analysis research.

Based on the background above regarding personal data protection laws, the researcher is interested in taking the title "Analysis of School Community Sentiment Against the Personal Data Protection Law Using the Support Vector Machine Method".

Related research

To support the research conducted, several studies were taken related to the application of the Support Vector Machine (SVM) method that had been carried out before, the following is a recapitulation table:

Table 1. Simplification of the List of Previous Research

No	Writer	Journal Title	Year	Conclusion	Best According to the Journal
4	(Naufal, Arifin, & Wirjawan, 2023)	Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial	2023	Algoritma SVM dan Random Forest memiliki performa yang terbaik dengan precision 82%, recall 83%, accuracy 83% dan precision 83%, recall 82%, accuracy 82%.	Nilai Sama
6	(Kurniawan et al., 2023).	Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter	2023	Tokopedia nilai accuracy NB 85.34%, dan SVM 86.82%, Shopee nilai accuracy NB 80.04%, dan SVM 80.91%. Lazada nilai accuracy NB 83.52%, dan SVM 88.93%	Support Vector Machine
7	(Nada, Soehardjoepri, & Atok, 2023)	Perbandingan Analisis Sentimen Mengenai BPJS pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM)	2023	klasifikasi SVM Kernel RBF, SVM Kernel Linear, dan Naïve Bayes Classifier masing-masing sebesar 97,1%, 92,5%, dan 86,7%.	Support Vector Machine
9	(Siregar, Ladayya, Albaqi, & Wardana, 2023)	Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter	2023	Metode klasifikasi yang menghasilkan prediksi terbaik pada data uji menggunakan kriteria F1-weighted average adalah SMOTE-SVM dengan nilai mencapai 60,45%.	Support Vector Machine
15	(Aldisa & Maulana, 2022)	Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM	2022	Skor AUC terbesar model SVM (75.40%), untuk presisi yang lebih akurat model Naive Bayes (83.81%).	Support Vector Machine
16	(Luthfanida, 2022)	Analisis Sentimen Data Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine (SVM) Tentang Presiden Jokowi 3 Periode	2022	hasil akurasi algoritma Naive Bayes sebesar 94,07% dan algoritma Support Vector Machine (SVM) sebesar 5,42%.	Support Vector Machine

17	(Pradana, Slamet, & Zukhronah, 2023)	Analisis Sentimen Kinerja Pemerintahan Menggunakan Algoritma NBC, KNN, dan SVM	2022	SVM kernel linier nilai akurasi 85,47%, nilai presisi 89,34%, nilai recall 90,34%, dan nilai F1-score 89,83%.	Support Vector Machine
21	Agustinus Ndruru (Ndruru, 2022)	Analisis Sentimen UU Cipta Kerja Melalui Omnibus Law Menggunakan Naïve Bayes Classifier (NBC) Dan Support Vector Machine (SVM)	2022	Klasifikasi NBC akurasi 95,6% dan 97,8%, nilai G-mean dan AUC 81,3% and 82,36%. Klasifikasi SVM akurasi sebesar 97,9% dan 99,3%, nilai G-mean dan AUC sebesar 97,35% and 97,38%.	Support Vector Machine
26	(Adrian, Putra, Rafialdy, & Rakhmawati, 2021)	Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB	2021	Akurasi algoritma Random Forest pada data yang di tes sebesar 0.578. Akurasi dari algoritma Support Vector Machine pada data yang di tes sebesar 0.557.	Random Forest
27	(Pamungkas & Kharisudin, 2021)	Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter	2021	Algoritma SVM memiliki akurasi yang lebih tinggi sebesar 90,01% pada SVM dengan kernel linear, 79,20% pada Naive Bayes dengan jumlah laplace adalah 1, dan 62,10% pada KNN dengan jumlah K adalah 20 dan menggunakan kernel optimal	Support Vector Machine
31	(Asshiddiqi & Lhaksmana, 2020)	Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI	2020	Hasil pengujian 80%:20% hasil nilai akurasi 87.45%, precison 87.72%,recall 91.74% dan F1-Score 89.69% pada Decision Tree, Support Vector Machine dengan 80%:20% hasil nilai akurasi 94.36%, precison 96.78%, recall 94.30% dan F1-Score 95.53%.	Support Vector Machine
32	(Fitri, 2020)	Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine	2020	Random Forest akurasi 97,16% nilai AUC 0,996, Support Vector Machine akurasi 96,01% nilai AUC 0,543 dan Naive Bayes akurasi 94,16% nilai AUC 0,999.	Random Forest
34	(Gunawan, Riana, Ardiansyah, Akbar, & Alfarizi, 2020)	Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023	2020	SVM akurasi 92.61% AUC 0,950, Naive Bayes akurasi 93,29% AUC 0,525, SVM Genetic Algorithm akurasi 93,03% AUC 0,869, Naive Bayes Genetic Algorithm akurasi 92,85% AUC 0,543	Naïve Bayes
42	(Pertiwi, 2019)	Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Mengguna-kan Algoritma Naïve Bayes, Neural Network, KNN dan SVM	2019	Algoritma k-NN akurasi yang lebih tinggi, akurasi= 90,76% dan AUC= 0,939; akurasi SVM= 89,03% dan AUC= 0,5; akurasi Naïve Bayes= 78,16% dan AUC= 0,567 dan Neural Network= 52,73% dan AUC=0,0.	k-Neural Network
43	(Riadi, Umar, & Aini, 2019)	Analisis Perbandingan Detection Traffic Anomaly dengan Metode Naive Bayes dan Support Vector Machine (SVM)	2019	nilai akurasi Naïve Bayes melalui data grafik Distributions dan Radviz memiliki nilai probabilitas 0.1 dan nilai probabilitas paling tinggi 0.8 hasil Support Vector Machine (SVM)	Support Vector Machine
47	(Najib, Irsyad, Qandi, & Rakhmawati, 2019)	Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter	2019	Metode Lexicon-based nilai akurasi 39% dan metode SVM sebesar 83%.	Support Vector Machine

Research Methods

The main point of study in this research is to discuss and analyze the concept of privacy protection for personal data as one of the constitutional rights of citizens and personal data privacy settings. This survey outlines the concept of protecting personal data privacy and protecting personal data privacy in the perspective of public sentiment, the extent to which knowledge of personal data is measured through self-assessment.

Respondents were also asked to fill out a number of knowledge questions about the Personal Data Protection Act (UU PDP).

There is a research hypothesis consisting of a research framework, data collection, and a questionnaire and its results. A research framework is a temporary description or explanation of a phenomenon in research and is our argument in formulating conclusions about what problems we will discuss in a study. the research framework that researchers use in the Analysis of School Community Sentiments towards the Personal Data Protection Law Using the Support Vector Machine Method, Case Study: SMK Negeri 1 Cikarang Selatan.

Data collection techniques or methods can be obtained using primary or secondary sources. Primary sources are data sources that provide data to data collectors, while secondary sources are data sources that indirectly provide data to data collectors, for example through documents or archives. In this study there were 4 data collection techniques that were carried out, namely observation, interviews, literature studies, and questionnaires.

One method of data collection carried out by the author is by conducting a questionnaire. This questionnaire was carried out with respondents from students/teachers/staff at SMK Negeri 1 Cikarang Selatan which was conducted randomly, consisting of five questions that had to be answered as a source of data obtained by the author in conducting this research. Below are the results of the complete questionnaire:

Table 2 Respondent Data Based on School Community Categories

Respondent Data	Number of Respondents	Percentage %
Instructor/Teacher	39	39%
School staff	8	8%
Student / Student	53	53%
Amount	100	100%

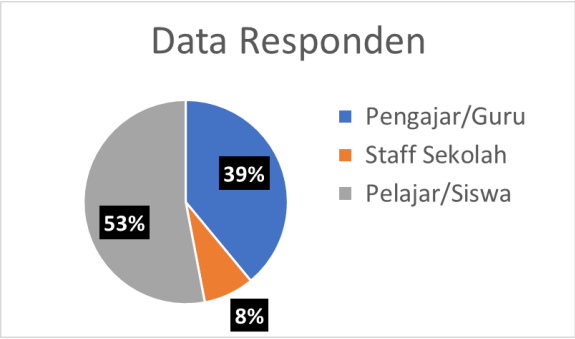


Figure 1 - Respondent Data by School Community Category
Source: Author

From the results of the published questionnaires, it was obtained that the data that had been selected were as many as 100 respondents consisting of 53 school student respondents, 39 teacher respondents and 8 school staff respondents. Then the respondent was given a number of questions and opinions regarding the Personal Data Protection Act, below are some of the results that can be summarized.

Table 3. Respondents have received socialization regarding the PDP Law

Ever received socialization of the PDP Law	Number of Respondents	Percentage %
Of	12	12%
No	88	88%
Amount	100	100%

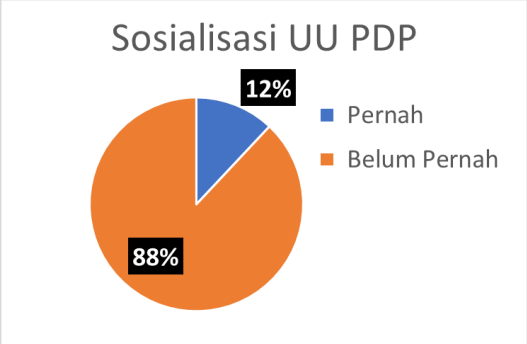


Figure 2 - Data on Socialization of the PDP Law
Source: Author

Table 4. Respondents consider the PDP Law to be effective

The PDP Act is effective protect personal data	Number of Respondents	Percentage %
Of	25	25%
No	12	12%
Maybe	63	63%
Amount	100	100%

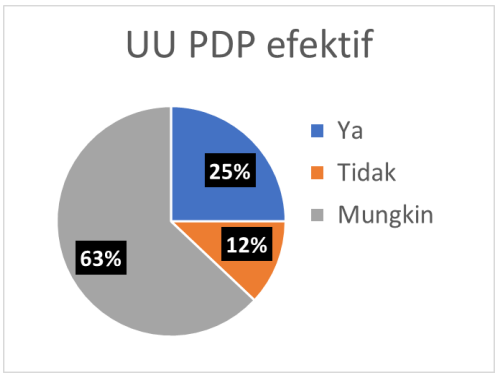


Figure 3 - Respondents' data consider the PDP law to be effective
Source: Author

Table 5. Respondents consider the PDP Law Useful

The PDP Act is useful to protect abuse	Number of Respondents	Percentage %
Of	56	56%

No	10	10%
Maybe	34	34%
Amount	100	100%

Based on the table above, it is known that of the 100 respondents who answered the questionnaire, as many as 56 respondents from the school community considered the PDP Law to be useful in protecting against misuse, so it can be concluded that 56% believed that the PDP Law could provide benefits to protect against data misuse that occurred. As many as 10 respondents or 10% of respondents still did not believe that the PDP Law would be useful in protecting data abuse and 34 respondents said they were still unsure about the PDP Law which might be useful in protecting data abuse. Detailed mapping results can be seen in the diagram below.

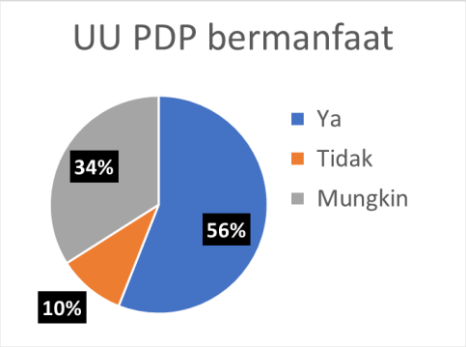


Figure 4 - Data Respondents consider the PDP Law useful
Source: Author

Results and Discussions

1.1.Data Acquisition Process

Research design is an arrangement regarding the requirements for data collection and analysis with the aim of linking a research objective with research procedures (Grenner & Martelli, 2018). The main goal in research design is to assist researchers in focusing their research so as to avoid data that has nothing to do with the research question. Research design is also ablueprint (blueprints) for the purpose of collecting, measuring and analyzing data. In the research design, it must contain several aspects such as:

1. Clear statement of the research problem
2. Procedures and techniques used
3. Population and sample to be studied
4. The method is suitable for research in analyzing data.

4.2 Preprocessing

The data obtained from the results questioner in the previous stage it cannot be used because it still has unstructured sentence forms so it needs to be processed preprocessing data. The questionnaire documents taken are still intact and have not been cleaned of components so that the processed data is still mixed with other characters which will make it difficult to perform data processing. The purpose of preprocessing is to remove noise, convert the data obtained so that it fits the needs either by increasing or decreasing the value of a data.

4.2.1 Case Folding

At the preprocessing stage there is a sectioncase folding with the aim of this section is to uniform the form of letters or words into lowercase and to remove punctuation in the review sentence. This is done so that words that have lowercase and uppercase letters are not detected or interpreted as having different meanings. There are no definite rules regarding the stages in text preprocessing. But ensuring better and consistent results if all stages are carried out. The more stages that are carried out, the more layers are peeled in the data. But one of the steps that need to be done isCase folding. This stage is the simplest preprocessing stage (Nugroho, 2019). Steps taken in the processCase Folding such as converting all letters in a sentence to lowercase which can improve classification success in terms of accuracy and help group terms that contain the same information (Uysal, 2014). The results of several processes in sentencesquestioner as below:

No	Persepsi mengenai UU PDP
1	uu pdp membantu melindungi data masyarakat dari penyalahgunaan data
2	selama kasus byorka dan pencurian data bank masih belum terproses, uu pdp hanya sebatas peraturan
3	semoga semakin banyak orang yang sadar dengan keamanan data
4	bagus untuk melindungi data privasi dari pencurian data
5	sebagai payung hukum supaya pelanggar dapat diberi hukuman dengan berat
6	sosialisasi masih minim, kalau ada kasus harus lapor ke lembaga mana
7	akhirnya indonesia juga memiliki peraturan mengenai data sehingga tidak tertinggal dari negara lain
8	data masih mungkin dapat dicuri kalau tidak ada tindakan atau kasus
9	tidak akan bermanfaat tanpa peranan masyarakat dalam menjaga data pribadi masing-masing
10	bermanfaat untuk menjaga data yang disimpan pada instansi supaya dijaga dengan baik
11	dasar peraturan yang baik dalam mencegah masalah pencurian data
12	lembaga penyelenggara belum ada yang menjamin keamanan data yang tersedia
13	pengamanan data pribadi supaya tidak disalahgunakan
14	perlu pencerahan dan sosialisasi di sekolah mengenai pentingnya menjaga data privasi
15	dasar hukum yang bagus supaya pelaku dapat berfikir ulang mengenai sanksi yang diberikan
....	....

Figure 5 - ResultsCase Folding data

Whereas in Figure 4.3 the process of removing punctuation marks such as commas (,) dashes (-) and other punctuation marks that are not needed in data processing is also known as the cleansing process, this process aims to help reduce noise/errors that occur in processing data later . These two stages are needed and very important to do in order to obtain valid data to be processed in the next process.

4.2.2 Tokenizing or Parsing

At the level oftokenizing or also known asParsing is a process of separating text into documents into pieces of words that influence each other and determine the syntactic structure of each available word. These words use a comma (,) as a separator. The following is an example of implementing tokenizing in a review sentence which can be seen in the table below:

D1	uu	pdp	membantu	melindungi	data	masyarakat	dari	penyalahgunaan	data
D2	selama	kasus	byorka	dan	pencurian	data	bank	masih	belum
	terproses	uu	pdp	hanya	sebatas	peraturan			
D3	semoga	semakin	banyak	orang	yang	sadar	dengan	keamanan	data
D4	bagus	untuk	melindungi	data	privasi	dari	pencurian	data	
D5	sebagai	payung	hukum	supaya	pelanggar	dapat	diberi	hukuman	dengan
	berat								
D6	sosialisasi	masih	minim	kalau	ada	kasus	harus	lapor	ke
	lembaga	mana							
D7	akhirnya	indonesia	juga	memiliki	peraturan	mengenai	data	sehingga	tidak
	tertinggal	dari	negara	lain					
D8	data	masih	mungkin	dapat	dicuri	kalau	tidak	ada	tindakan
	atau	kasus							
D9	tidak	akan	bermanfaat	tanpa	peranan	masyarakat	dalam	menjaga	data
	pribadi	masing	masing						
D10	bermanfaat	untuk	menjaga	data	yang	disimpan	pada	instansi	supaya
	dijaga	dengan	baik						
D11	dasar	peraturan	yang	baik	dalam	mencegah	masalah	pencurian	data
D12	lembaga	penyelenggara	belum	ada	yang	menjamin	keamanan	data	yang
	tersedia								
D13	pengamanan	data	pribadi	supaya	tidak	disalahgunakan			
D14	perlu	pencerahan	dan	sosialisasi	di	sekolah	mengenai	pentingnya	menjaga
	data	privasi							
D15	dasar	hukum	yang	bagus	supaya	pelaku	dapat	berfikir	ulang
	mengenai	sanksi	yang	diberikan					

Figure 6 -ResultsTokenizing data
Source: Author

4.2.3 Filtering / Stopword Removal

At this stage, the process of removing the words contained in the sentences that have been obtained by usingstopword removal. Stopword Removal used aims to eliminate words that have no influence on the sentence by not reducing the information or meaning of the sentence itself. Stagestopword removal saves a lot of memory space and processing time and does not damage the effectiveness of the information retrieved (Jha, 2016). Process donestopword removal then the resulting information is more focused on important information and deletion is done totext which has low level information. ProcessStopword removal will reduce the size of the dataset thus will reduce the time of the training process because the amount of data used will be reduced. When stagestopword removal removed it will affect the data processing time and the number of data sets is also more than when usingstopword removal (Khanna, 2021). As for the processStopword Removal can be done as follows:

D1	uu	pdp	membantu	melindungi	data	masyarakat		penyalahgunaan	data
D2	selama	kasus	byorka		pencurian	data	bank		belum
	terproses	uu	pdp		sebatas	peraturan			
D3	semoga	semakin	banyak	orang		sadar		keamanan	data
D4	bagus		melindungi	data	privasi		pencurian	data	
D5	sebagai	payung	hukum		pelanggar		diberi	hukuman	
	berat								
D6	sosialisasi	masih	minim			kasus	harus	lapor	
	lembaga	mana							
D7	akhirnya	indonesia		memiliki	peraturan	mengenai	data		tidak
	tertinggal		negara	lain					
D8	data	masih	mungkin		dicuri		tidak		tindakan
		kasus							
D9	tidak		bermanfaat	tanpa	peranan	masyarakat		menjaga	data
	pribadi	masing	masing						
D10	bermanfaat		menjaga	data		disimpan		instansi	
	dijaga		baik						
D11	dasar	peraturan		baik		mencegah	masalah	pencurian	data
D12	lembaga	penyelenggara	belum			menjamin	keamanan	data	
	tersedia								
D13	pengamanan	data	pribadi		tidak	disalahgunakan			
D14	perlu	pencerahan		sosialisasi		sekolah	mengenai	pentingnya	menjaga
	data	privasi							
D15	dasar	hukum		bagus		pelaku		berfikir	ulang
	mengenai	sanksi		diberikan					

Figure 7 - ProcessFiltering data
Source: Author

From table 4.5 it is known that there are several missing words, because the function of this data filtering process is to remove words that have no meaning or words that are considered unimportant such as conjunctions (which, in, to, yes). This stage can be said to be quite complicated if there are several words that are not in accordance with the standard language of the large Indonesian dictionary. In addition, if there are connecting words that blend with other words, it can also complicate the filtering process.

4.2.4 Labeling

After doing the stepspreprocessing, then the next stage is the labeling stage. In carrying out the labeling process can be done in several ways, namely manually and usingLexicon-based. The manual labeling process is carried out based on individual domain knowledge and understanding of the language. However, this process will take quite a lot of time (Verma, 2018). Another method that is often used isLexicon-based. MethodLexicon-based This is done by creating a dictionaryLexicon. Where the dictionary will be used to identify whether the word contains an opinion or not.Lexicon is a collection of known and collected sentiment words (Desai, 2016). The labeling process requires a dictionary orlexicon which contains words that contain sentiments calledsentiment dictionaries (Buntoro, 2014). The results of the labeling process can be seen in the table below this.

Table 6. Labeling Results

Labeling Analysis	Analysis Datasets
Positive	496
Neutral	319
Negative	72

Amount	887

Based on the table above, it can be explained that the positive sentiment is 496 points, the negative sentiment is 72 points and the neutral sentiment class is 319 points. Where the sentiment class is a statement of school community sentiment regarding the personal data protection law that has been carried outquestioner previously.

4.3 Results of Analysis of the Dataset

Data successfully obtained fromquestioner towards the school community and then processed through the text analysis level. At the level oftext analysis, the text obtained is processed usingmachine learning andknowledge based. separation of each text with notation of words into positive, negative, and neutral groups, according to predetermined data. After that the data is displayed in the graph and analysis of the data obtained.

The results of the test were carried out on a dataset of school community opinion on the 2021 Personal Data Protection Act (UU PDP) obtained through a questionnaire of 1128 lines of data which were then processed into 887 data which have been checked so that the final results can be seen in Figure 4.8 below.

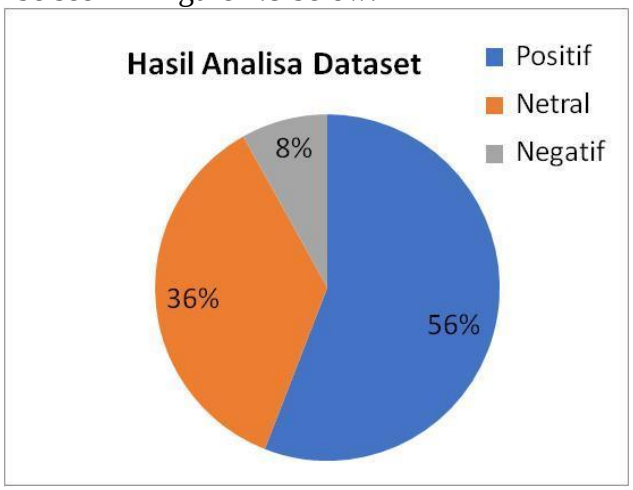


Figure 8 - Results of Dataset Analysis
Source: Author

Datasets that are applied manually weighted or labeled data predict that the sentiments conveyed by the school community in the questionnaires conducted, obtained 56% or as many as 496 words out of 887 Sentiments conveyed show a positive response to the Personal Data Protection Act (UU PDP). Results were also obtained by 36% or as many as 319 words from 887 Sentiments which showed a neutral response and 8% or as many as 72 words from 887 Sentiments which showed a negative response. The total distribution of the entire dataset can be seen in the image below.

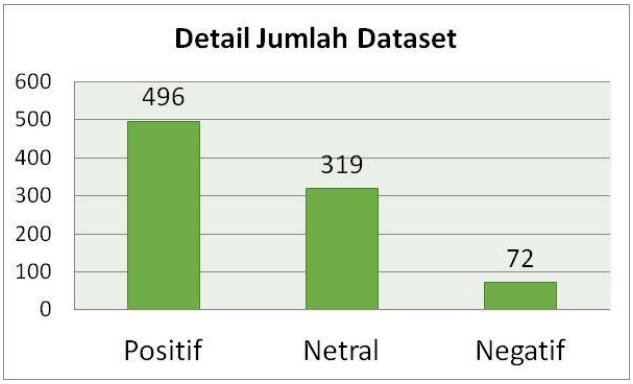
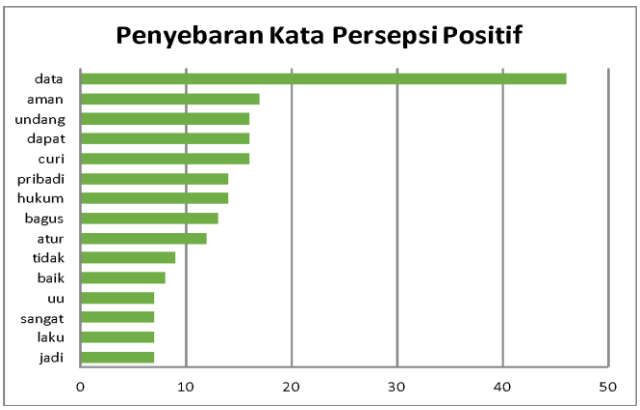


Figure 9 - Detailed Amount of Dataset Analysis
Source: Author

4.3.1 Positive Sentiment

In the visualization and association stages of positive reviews, data is collected based on the results of the manual labeling process that was carried out in the previous stage. Based on these data the spread of words on positive sentiment is shown in the figure below.



Picture 10- The spread of words with positive sentiments
Source: Author

The graph in Figure 10 shows the results of word distribution in the positive sentiment class. Where the word that appears most often is the word data which appears as 46 times. The next visualization is done using wordcloud. The following is a visualization using wordcloud that has been done:

Figure 11- WordCloud of positive Sentiment words

Based on Figure 11 the results from wordcloud on positive sentiment. Visualization using wordcloud is carried out with the aim of being able to provide a clearer picture of the words that often appear in positive sentiment. If the word that appears is getting bigger, it indicates that the word appears more often or the higher the frequency of the word appearing in positive sentiment. Figure 4.11 shows the words that have the highest frequency are data, safe, invite, can and others.

4.3.2 Neutral Sentiment

The visualization and association review stages were also carried out on data sentiment which was included in the neutral category based on the results of the manual labeling process that had been carried out in the previous stage. Based on these data, the distribution of words on neutral sentiment is obtained as shown in the figure below.

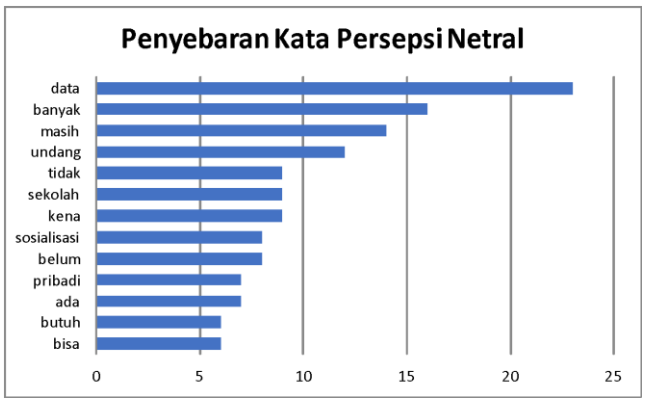


Figure 12- Spread of neutral sentiment words
Source: Author

The graph in Figure 12 shows the results of word distribution in the neutral sentiment class. Where the word that appears most often is also the word data which appears as 23 times. The next visualization is done using wordcloud. The following is a visualization using wordcloud that has been done:

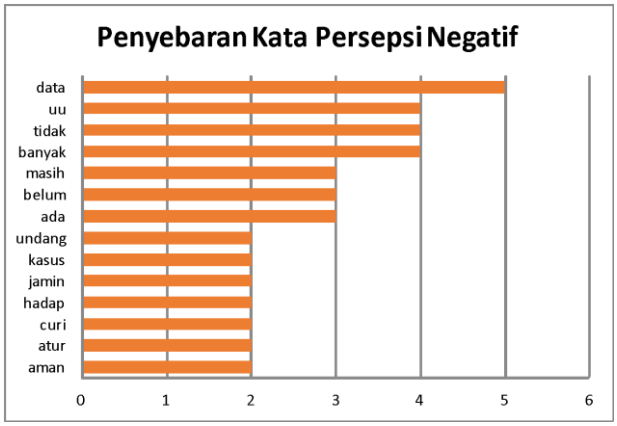
Figure 13-WordCloud neutral sentiment
Source: Author

Based on Figure 13 shows the results ofwordcloud neutral sentiment. Visualization usingwordcloud carried out with the aim of providing a clearer picture of the words that often appear in neutral sentiments. If the word that appears is getting

bigger, it indicates that the word appears more often or the higher the frequency of the word appearing in neutral sentiment. Figure 4.13 shows the words that have the highest frequency are data, many, still, invite, and others.

4.3.3 Negative Sentiment

In addition to positive and neutral sentiments, the visualization and association review stages were also carried out on data sentiment which was included in the negative category based on the results of the manual labeling process that had been carried out in the previous stage. Based on these data, the distribution of words on negative sentiment is shown in Figure 14.



Picture 14- The spread of words with negative sentiments
Source: Author

The graph in Figure 14 shows the results of word distribution in the negative sentiment class. Where the word that appears most often is also the data word that appears as five times. Further visualization is carried out usingwordcloud. The following is a visualization using wordcloud that has been done:

Figure 15-WordCloud negative sentiment
Source: Author

Based on Figure 15 shows the results of word cloud on negative sentiment. Visualization using word cloud This is done with the aim of providing a clearer picture of the words that often appear in negative sentiment. The bigger the word that appears, it means that the word appears more often or the higher the frequency of the word appears

in negative sentiment. In Figure 4.15 the words that have the highest frequency are data, uu, no, many and others.

4.4 Results Discussion

Personal data and confidentiality (privacy) are one unit that cannot be split into two parts. This is because personal data is interrelated with privacy, when we talk about personal data owned by someone, indirectly we are also talking about the privacy of that person which must be protected and respected. Privacy is another term that is then used by developed countries relating to personal data as a right that must be protected, namely the right of a person not to be disturbed by his private life.

4.4.1 Classification withSupport Vector Machine

Table 7. Distribution of training data and test data

Data Comparison		Amount of data	
Training Data	Test Data	Training Data	Test Data
90%	10%	90	10
80%	20%	80	20
70%	30%	70	30
60%	40%	60	40
50%	50%	50	50

From table 4.2 The distribution of training data and test data from the results of data acquisition and preprocessing that was carried out previously obtained as many as 100 data which were divided into 3 sentiment classes, namely positive, neutral and negative. The data is then divided into 2 categories, namely training data and test data before being classified using the methodSupport Vector Machine This test was carried out with 5 different distribution of training data and test data.

4.4.2 Experiment with 90% Training Data and 10% Test Data

Table 8. Experimental training data 90%

Actual Data	Predictive Data			Precision	Recall	F-Measure	Class
	Positive	Neutral	Negative				
Positive	76	0	0	0,8444	1,0000	0,9157	Positive
Neutral	10	0	0				Neutral
Negative	4	0	0				Negative
	90	0	0	84,44%	100,00%	91,57%	
Accuracy	84,44%	Total	90				

In table 4.4 it can be seen the results of data calculationsaccuracy, precision, recall andF-Measure with a data set that has been divided by the provisions of 90% training data and 10% test data with a total data processing of 90 data with 76 correctly predicted positive data results, 0 positive predicted neutral data, 0 positive predicted negative data

0, then neutral data with a positive prediction of 10, neutral data with a correct prediction of 0, neutral data with a negative prediction of 0 and negative data with a positive prediction of 4.

To ensure the correctness of the data results that have been processed and calculatedaccuracy, precision, recall andF-Measure can be checked by calculating the formulamatrix confusion with the 3x3 class so that the same results are obtained, namely as follows

Positive Class :

$$\text{pre} - \text{positive} = \frac{\text{TP}}{\text{TP} + \text{OP} + \text{FP}} = \frac{76}{76 + 10 + 4} = \frac{76}{80} = 0.844$$

$$\text{re} - \text{positive} = \frac{\text{TP}}{\text{TP} + \text{PO} + \text{FP}} = \frac{76}{76 + 0 + 0} = \frac{76}{76} = 1$$

$$\text{F1} - \text{positive} = 2 \times \frac{\text{pre} - \text{positive} \times \text{re} - \text{positive}}{\text{pre} - \text{positive} + \text{re} - \text{positive}} = 2 \times \frac{0.844 \times 1}{0.844 + 1} = 0.9157$$

Neutral Class :

$$\text{pre} - \text{neutral} = \frac{\text{TO}}{\text{TO} + \text{PO} + \text{NO}} = \frac{0}{0 + 0 + 0} = \frac{0}{0} = \text{N\#A}$$

$$\text{re} - \text{netral} = \frac{\text{TO}}{\text{TO} + \text{OP} + \text{ON}} = \frac{0}{0 + 10 + 0} = \frac{0}{10} = \text{N\#A}$$

$$\text{F1} - \text{netral} = 2 \times \frac{\text{pre} - \text{neutral} \times \text{re} - \text{neutral}}{\text{pre} - \text{netral} + \text{re} - \text{netral}} = 2 \times \frac{\text{N\#A} \times \text{N\#A}}{\text{N\#A} + \text{N\#A}} = \text{N\#A}$$

Negative Class :

$$\text{pre} - \text{negative} = \frac{\text{TN}}{\text{TN} + \text{FP} + \text{ON}} = \frac{0}{0 + 0 + 0} = \frac{0}{0} = \text{N\#A}$$

$$\text{re} - \text{negative} = \frac{\text{TN}}{\text{TN} + \text{FN} + \text{NO}} = \frac{0}{0 + 4 + 0} = \frac{0}{4} = \text{N\#A}$$

$$\text{F1} - \text{negative} = 2 \times \frac{\text{pre} - \text{negative} \times \text{re} - \text{negative}}{\text{pre} - \text{negative} + \text{re} - \text{negative}} = 2 \times \frac{\text{N\#A} \times \text{N\#A}}{\text{N\#A} + \text{N\#A}} = \text{N\#A}$$

The conclusion obtained from the calculation resultsaccuracy, precision, recall andF-Measure it can be obtained the average result of the classificationSupport Vector Machine (SVM) as shown in the following table below:

Table 9. Results of 90% training data

SVM value	Yield Value
Training Data : 90 %	90
Test Data : 10 %	10
Accuracy Classification	84.44%
Average Precision	84.44%
Average Recall	100%
Average F-Measure	91.57%

4.4.3 Analysis of Experimental Data Results

After analyzing the results with five (5) trials with differences in training data and test data, in accordance with the initial objectives of the study, the sentiment analysis of the school community used Support Vector Machine quite ideal to be used as a data observation method. The use of parameters is seen from Accuracy, Precision, Recall and F-Measure to get the most optimal analysis results. For this reason, the results of the values of these parameters can be summarized as in the table below:

Table 10. Classification Results with SVM

SVM value	Result Value (Train Data % - Test Data %)					Average
	90%-10%	80%-20%	70%-30%	60%-40%	50%-50%	
Training Data	90	80	70	60	50	
Test Data	10	20	30	40	50	
Accuracy Classification	84,44%	85,00%	85,71%	86,67%	88,00%	85,97%
Average Precision	84,44%	85,00%	85,71%	86,67%	88,00%	85,97%
Average Recall	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Average F-Measure	91,57%	91,89%	92,31%	92,86%	93,62%	92,45%

It can be concluded from the results of the calculation value using the method Support Vector Machine for the highest accuracy results obtained in the experiment of training data and test data 50% - 50% which is equal to 88.0% in the second place the experiment on training data and test data 60% - 40% which is equal to 86.67% then in the third order the accuracy of the experiment on training data and data test 70% - 30%, which is equal to 85.71%, then the fourth position is obtained in the experiment of training data and test data of 80% - 20%, which is equal to 85.00% and the lowest result is in the experiment of training data and test data, which is 90% - 10%, which is equal to 84.44% .

As for value precision the average of the five trials was 85.97% with the highest results obtained in experiments with training data and test data 50% - 50%, namely 88.00% in the second place experiments on training data and test data 60% - 40%, namely 86.67% then in the order of three values precision experiments on training data and test data 70% - 30%, which is equal to 85.71%, then the fourth position is obtained in the trial data, training data and test data 80% - 20%, which is equal to 85.00% and the lowest result is in the experiment, training data and test data 90% - 10% that is equal to 84.44%.

for value recall from each experiment the training data and test data obtained an average yield of 100% with detailed results of all experiments that is equal to 100%. After that the last test to determine the value of F-Measure, after doing the calculations, the experimental results obtained an average value of 92.45% for the results F-Measure the highest was obtained in the experiment of training data and test data of 50% - 50%, namely 93.62%, in the second place the experiment on training data and test data was 60% - 40%, namely 92.86%, then in the third place the results F-Measure experiments on training data and test data 70% - 30%, which is equal to 92.31%, then the fourth position is obtained in the experiment, training data and test data 80% - 20%, which is equal to 91.89% and the lowest result is in the experiment, training data and test data 90% - 10% that is equal to 91.57%.

4.4.8 Comparative Analysis of Trial Data

As proof that the results of data testing using the SVM method (Support Vector Machine) is a good method for testing the data, so data testing is also carried out using other methods such as Naive Bayes and K-NN with the same experiment five times with training data and test data with different percentages ranging from 90% - 10 %, 80% - 20%, 70% - 30%, 60% - 40% and training data - test data of 50% - 50% the same as done using the methodSupport Vector Machine The following below is the result of calculations that have been carried out on the school community's perception data of the Personal Data Protection Law

Table 11. Classification Results with Naïve Bayes

Naive Bayes Value	Result Value (Train Data % - Test Data %)					Average
	90%-10%	80%-20%	70%-30%	60%-40%	50%-50%	
Training Data	90	80	70	60	50	
Test Data	10	20	30	40	50	
Accuracy Classification	84,44%	85,00%	82,86%	85,00%	88,00%	85,06%
Average Precision	57,96%	62,61%	60,51%	60,53%	70,74%	62,47%
Average Recall	49,91%	48,53%	60,00%	64,74%	73,86%	59,41%
Average F-Measure	52,24%	52,16%	59,89%	62,54%	72,25%	59,82%

From the table above, it can be seen the results for Accuracy, Precision,Recall and F-Measure obtained by training data criteria - test data 50% - 50% with each accuracy value of 88.00%, precision value of 70.74%, valuerecall of 73.86% and valueF-Measure by 72.25%. for the value of accuracy the results are the same as using the methodSupport Vector Machine but for other category values it is still greater using the methodSupport Vector Machine compared to the Naive Bayes method.

Table 12. Classification Results with KNN

KNN value	Result Value (Train Data % - Test Data %)					Average
	90%-10%	80%-20%	70%-30%	60%-40%	50%-50%	
Training Data	90	80	70	60	50	
Test Data	10	20	30	40	50	
Accuracy Classification	80,00%	77,50%	78,57%	80,00%	82,00%	79,61%
Average Precision	84,71%	83,78%	83,33%	85,71%	87,23%	84,95%
Average Recall	94,74%	93,94%	94,83%	96,00%	97,62%	95,42%
Average F-Measure	89,44%	88,57%	88,71%	90,57%	92,13%	89,88%

The results of calculations using the KNN method also obtained the highest value in the category of training data - test data 50% - 50% with an accuracy of 82.00%, a precision value of 87.23%, a recall value of 97.62% and an F-Measure value of 92.13%.

Conclusion

Based on the results of research and data trials that have been carried out and processed into sentiment analysis using the Support Vector Machine method for the Personal Data Protection Act with the correspondents of the school community at SMK Negeri 1 Cikarang Selatan, the conclusions that can be drawn are as follows: a. Basically, the school community knows about the Personal Data Protection Law that has been set by the government, but based on the results of the new questioner, only 38% know about this rule while the other 62% still do not know much about this rule. b. Based on the results of sentiments made on opinions or comments obtained through questionnaires to the school community regarding the Personal Data Protection Act, it can be concluded that the school community views the PDP law positively. This can be proven by the results obtained, namely 56% or as many as 496 of 887 words. Sentiments submitted showed a positive response to the Personal Data Protection Act (UU PDP). while the results were 36% or 319 of 887 sentiment words showing a neutral response and 8% or 72 of 887 sentiment words showing a negative response. c. Tests using the Support Vector Machine (SVM) method have been carried out five (5) trials using different variations of training data and test data resulting in an average accuracy rate of 85.97% with the highest results on training data and test data of 50% - 50%, which is equal to 88.00% and the lowest result is in the experiment of training data and test data of 90% - 10%, which is equal to 84.44%.

References

Adrian, Muhammad Rivza, Putra, Muhammad Papuandivitama, Rafialdy, Muhammad Hilman, & Rakhmawati, Nur Aini. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB. Jurnal Informatika Upgris, 7(1).

Aldisa, Rima Tamara, & Maulana, Pandu. (2022). Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM. Building of Informatics, Technology and Science (BITS), 4(1), 106–109.

Asshiddiqi, Muhammad Fadli, & Lhaksmana, Kemas Muslim. (2020). Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI. EProceedings of Engineering, 7(3).

Fitri, Evita. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. Jurnal Transformatika, 18(1), 71–80.

Gunawan, Deni, Riana, Dwiza, Ardiansyah, Dian, Akbar, Fajar, & Alfarizi, Salman. (2020). Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023. V (1), 135–138.

Kurniawan, Indra, Hananto, April Lia, Hilabi, Shofa Shofia, Hananto, Agustia, Priyatna, Bayu, & Rahman, Aviv Yuniar. (2023). Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter. JATISI (Jurnal Teknik Informatika Dan Sistem Informasi), 10(1), 731–740.

Luthfanida, Luthfanida. (2022). Analisis Sentimen Data Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine (SVM) Tentang Presiden Jokowi 3 Periode. Djtechno: Jurnal Teknologi Informasi, 3(1), 5–11.

Nada, Diva Durrotun, Soehardjoepri, Soehardjoepri, & Atok, R. Mohamad. (2023).

Perbandingan Analisis Sentimen Mengenai BPJS pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM). Jurnal Sains Dan Seni ITS, 11(6), D480–D485.

Najib, Ahmad Choirun, Irsyad, Akhmad, Qandi, Ghiffari Assamar, & Rakhmawati, Nur Aini. (2019). Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter. Fountain of Informatics Journal, 4(2), 41–48.

Naufal, Mohammad Farid, Arifin, Theofilus, & Wirjawan, Hans. (2023). Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial. Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika), 8(1), 82–90.

Ndruru, Agustinus. (2022). Analisis Sentimen UU Cipta Kerja Melalui Omnibus Law Menggunakan Naive Bayes Classifier (NBC) Dan Support Vector Machine (SVM). Pelita Informatika: Informasi Dan Informatika, 10(3), 85–90.

Pamungkas, Fajar Sodik, & Kharisudin, Iqbal. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. PRISMA, Prosiding Seminar Nasional Matematika, 4, 628–634.

Pertiwi, Melisa Winda. (2019). Analisis sentimen opini publik mengenai sarana dan transportasi mudik tahun 2019 pada twitter menggunakan algoritma naïve bayes, neural network, KNN dan SVM. Inti Nusa Mandiri, 14(1), 27–32.

Pradana, Hizkia Yotant, Slamet, Isnandar, & Zukhronah, Etik. (2023). Analisis Sentimen Kinerja Pemerintahan Menggunakan Algoritma NBC, KNN, dan SVM. Prosiding Simposium Nasional Multidisiplin (SinaMu), 4, 114–121.

Riadi, Imam, Umar, Rusydi, & Aini, Fadhilah Dhinur. (2019). Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm). ILKOM Jurnal Ilmiah, 11(1), 17–24.

Siregar, Dania, Ladayya, Faroh, Albaqi, Naufal Zhafran, & Wardana, Bintang Mahesa. (2023). Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter. Jurnal Statistika Dan Aplikasinya, 7(1), 93–104.