JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 1 AGUSTUS 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4091

SENTIMENT ANALYSIS OF CONTENT PERMENKOMINFO NO.5 OF 2020 USING A CLASSIFICATION ALGORITHM

Mohammad Amada[1];Munawar[2*];Marzuki Pilliang[3]

Master of Computer Science[1,2,3]
Esa Unggul University
https://www.esaunggul.ac.id/
arabamada@gmail.com, moenawar@gmail.com, marzuki.pilliang@ieee.org

Abstract—*This study aims to evaluate the impact of the policy issued by the Minister of Communication and Information Technology (PERMENKOMINFO No.5 of 2020) on the public's ability to access content through Private Scope Electronic System Providers (PSE). The study uses sentiment analysis and data classification methods to analyze the content of PERMENKOMINFO No.5 of 2020 and provides results on the accuracy of sentiment prediction. The study results show that the data classification method in sentiment analysis can provide accurate results in predicting the sentiment toward the content of PERMENKOMINFO No.5 of 2020. The study also highlights the need for improvement and better policy to ensure the interests of the public in accessing online information. The negative sentiment of 80.34% obtained through sentiment analysis provides important contributions for policy evaluation and feedback for improvement. This study provides valuable insights into the public's sentiment towards the PERMENKOMINFO No.5 of 2020 policy and its impact on their ability to access content. It also contributes to understanding the legal uncertainty in accessing content and reinforces the case for better policy to ensure the interests of the public.*

Keywords: *Sentiment analysis, PERMENKOMINFO No.5 of 2020, Content Regulation.*

Intisari— Penelitian ini bertujuan untuk mengevaluasi dampak atas kebijakan yang diterbitkan oleh Menteri Komunikasi dan Teknologi Informasi (PERMENKOMINFO No.5 Tahun 2020) pada kemampuan masyarakat untuk mengakses konten melalui Penyelenggara Sistem Elektronik (PSE) Lingkup Pribadi. Studi ini menggunakan metode analisis sentimen dan klasifikasi data untuk menganalisis isi PERMENKOMINFO No.5 Tahun 2020 dan memberikan hasil tentang akurasi prediksi sentimen. Hasil studi menunjukkan bahwa metode klasifikasi data dalam analisis sentimen mampu memberikan hasil yang akurat dalam memprediksi sentimen terhadap isi PERMENKOMINFO No.5 Tahun 2020. Studi ini juga menyoroti kebutuhan untuk peningkatan dan kebijakan yang lebih baik untuk memastikan kepentingan masyarakat dalam mengakses informasi online. Sentimen yang negatif sebesar 80.34% yang didapatkan melalui analisis sentimen memberikan kontribusi penting bagi evaluasi kebijakan dan memberikan masukan untuk perbaikan. Studi ini memberikan wawasan penting tentang sentimen masyarakat terhadap kebijakan PERMENKOMINFO No.5 Tahun 2020 dan dampaknya pada kemampuan masyarakat untuk mengakses konten. Studi ini juga berkontribusi pada pemahaman tentang ketidakpastian hukum dalam mengakses konten dan memperkuat case untuk kebijakan yang lebih baik untuk memastikan kepentingan masyarakat.

Kata Kunci: *Analisis sentimen, PERMENKOMINFO No.5 Tahun 2020, Pengaturan Konten .*

INTRODUCTION

The Minister of Communication and Information Technology issued Ministerial Regulation Number 5 of 2020, which limits the activities of Private Scope Electronic System Providers (PSE Lingkup Privat). Through Press Release No. 296 in 2022, the Ministry of Communication and Information Technology set a deadline until July 27, 2022, for unregistered PSEs to process their registration, or they will be subject to temporary access termination sanctions. On July 30, 2022, unregistered PSEs were subjected to temporary service termination sanctions by the Ministry of Communication and Information Technology, making their services unavailable to users. This access closure triggered public responses on various social media platforms, including Twitter, and created the hashtag #BlokirKominfo, which shows various responses to the legal uncertainty for the public in accessing content after the imposition of sanctions on PSE by the Ministry of Communication and Information Technology.

Gap that creates legal uncertainty for the public in accessing content can be found in the administrative sanctions imposed by the Minister on PSEs that do not comply with the registration provisions stated in Article 2 and 6. Administrative sanctions such as access termination to electronic systems and even the revocation of electronic system provider registration can create legal uncertainty for the public in accessing content through the electronic system.

To understand the rejection that arose from the policy outlined in PERMENKOMINFO No. 5 of 2020 related to legal uncertainty in accessing content, it is necessary to analyze the content of the regulation. Sentiment analysis processes unstructured words found in a word or sentence to predict the expressive value in positive, negative, or neutral responses [1]. Sentiment analysis seeks to determine the polarity of text by measuring its level of subjectivity, which can result in a positive, neutral, or negative value [2]. The sentiment analysis research on the content of PERMENKOMINFO No.5 of the Year 2020 aims to identify public sentiment towards the policies outlined in it.

In sentiment analysis, several classification methods can be used. One is Naive Bayes, which operates based on probability and Bayesian theorem [3]. In this method, the assumption is that the presence or absence of an attribute is not related to other attributes. It calculates the probability of each previously determined data feature, calculates the likelihood value by multiplying the value of each probability, and then predicts the data based on the previously determined label [4]. In addition, Support Vector Machine (SVM) can also be used as a classification method. SVM is a hypothetical spatial learning system with a linear function in multidimensional features trained using optimization theory [5]. Finally, the K-Nearest Neighbor (K-NN) method divides new data based on the distance between new data and the nearest neighbors [6]. Several studies have utilized sentiment analysis and data classification as their research methods to analyze and understand public sentiment toward various issues.

Yoga Vikriansyah Wijaya, et al. [5] performed sentiment analysis on Twitter social media using the Support Vector Machine classification algorithm on policies related to the ITE Law. The test data accuracy was 84%, the recall value was 65%, and the f1-score was 71% for each sentiment class, with the majority indicating that 74.10% of the public is against the ITE Law.

Ahmad Syaifuddin and Mohammad Muslimin [7] analyzed opinion sentiment on Twitter using the Lexicon Based method on the Kominfo PSE policy. The sentiment analysis results were 80.1% negative towards the opinions expressed by the public on social media Twitter in a dataset of 3,300 tweets, leading to the conclusion that the public opposes the PSE policy.

The research conducted by Yoga Vikriansyah Wijaya et al., Ahmad Syaifuddin, and Mohammad Muslimin focused on public sentiment analysis regarding the PSE policy and ITE Law through the social media platform Twitter. However, this research has not demonstrated an in-depth understanding of the legal uncertainty for the public in accessing content. Therefore, sentiment analysis of the PSE policy and ITE Law content is necessary to understand the gap in legal uncertainty for the public in accessing content. This is a critical issue for understanding the implications of the PSE policy and ITE Law for the public.

The following research will focus on sentiment analysis of the content of Ministerial Regulation Number 5 of 2020. Sentiment analysis will help to understand how the public perceives the content of the policy and whether there are legal issues for the public in accessing content. This research will help researchers obtain objective and detailed information about the gap.

MATERIALS AND METHODS

The research that will be conducted combines a text analysis approach, namely Text Mining and sentiment analysis technique. Text Mining is searching for patterns in the text to extract useful information, while sentiment analysis understands a text's polarity and categorizes it as positive, negative, or neutral [8], [9]. In this research, the analysis process will go through several stages, such as dataset retrieval, Text Preprocessing, sentiment labeling with InSet Lexicon, word weighting with TF-IDF, and classification using Support Vector Machine, Naïve Bayes Classifier, and K-Nearest Neighbors. The research steps will be explained in Figure 1.

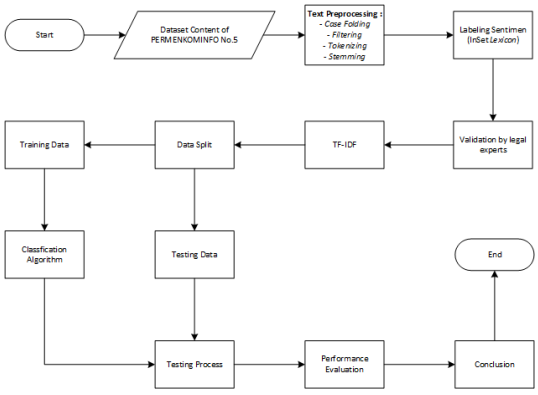

Figure 1. Research Procedure

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 8. NO. 1 AGUSTUS 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4091

Data Collection

In this research, the dataset used is derived from the content of Regulation of the Minister of Communication and Information Technology No. 5 of 2020. The dataset can be downloaded through the official website of the Legal Documentation and Information Network of the Ministry of Communication and Information Technology (JDIH Kominfo) at https://jdih.kominfo.go.id/produk_hukum/view/id/759/t/peraturan+menteri+komunikasi+dan+informatika+nomor+5+tahun+2020. The dataset consists of 178 section, starting from the opening until the end. In this study, sentiment analysis will be focused on articles 7 and 8, which contain Administrative Sanctions and Normalization and consist of 8 sections.

Text Preprocessing

Text Preprocessing is a process that is applied to change and clean text to be structured [8], [10] so that the text becomes quality when analyzed. At this stage, the dataset is processed to clean up words, characters, or symbols without value or influence on the analysis process and normalize words to become standard words/language. This process is divided into several stages, and the following steps are carried out :

1. Case Folding

Case Folding is a procedure used to equate the shape of letters in data into a lowercase form [11]. The output of the Case Folding process is described in table 1.

Table 1. Application of Case Folding in Datasets

Before	After
Dalam hal sanksi administratif yang diberikan kepada PSE Lingkup Privat sebagaimana dimaksud pada ayat (1) adalah Pemutusan Akses terhadap Sistem Elektronik (access blocking), Menteri melakukan Normalisasi berdasarkan pengajuan rekomendasi oleh Kementerian atau Lembaga atas dasar layanan PSE lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan.	dalam hal sanksi administratif yang diberikan kepada pse lingkup privat sebagaimana dimaksud pada ayat (1) adalah pemutusan akses terhadap sistem elektronik (access blocking), menteri melakukan normalisasi berdasarkan pengajuan rekomendasi oleh kementerian atau lembaga atas dasar layanan pse lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan.

2. Filtering

Filtering removes words/symbols considered to have no meaning [12]. In the Filtering stage, the text that the Case Folding process has carried out will be cleaned of punctuation, numbers, hashtags, mentions (user embeds), slang conversions, and correction of wrong/typo words (slang words) from the dataset used. The output of Filtering is described in table 2.

Table 2. Application of Filtering in Datasets

Before	After
dalam hal sanksi administratif yang diberikan kepada pse lingkup privat sebagaimana dimaksud pada ayat (1) adalah pemutusan akses terhadap sistem elektronik (access blocking), menteri melakukan normalisasi berdasarkan pengajuan rekomendasi oleh kementerian atau lembaga atas dasar layanan pse lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan.	dalam hal sanksi administratif yang diberikan kepada pse lingkup privat sebagaimana dimaksud pada ayat adalah pemutusan akses terhadap sistem elektronik access blocking menteri melakukan normalisasi berdasarkan pengajuan rekomendasi oleh kementerian atau lembaga atas dasar layanan pse lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan

3. Tokenizing

Tokenizing is splitting each existing sentence into words or tokens [13]. In the Tokenizing process, the entire dataset will be broken down per word so that each word is separate and does not become a complete sentence. This process makes it easier for the following process, namely the Stemming process. The output of Tokenizing is described in table 3.

CC BY NC

Table 3. Application of Tokenizing in Datasets.

Before	After
dalam hal sanksi administratif yang diberikan kepada pse lingkup privat sebagaimana dimaksud pada ayat adalah pemutusan akses terhadap sistem elektronik access blocking menteri melakukan normalisasi berdasarkan pengajuan rekomendasi oleh kementerian atau lembaga atas dasar layanan pse lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan	['dalam', 'hal', 'sanksi', 'administratif', 'yang', 'diberikan', 'kepada', 'pse', 'lingkup', 'privat', 'sebagaimana', 'dimaksud', 'pada', 'ayat', 'adalah', 'pemutusan', 'akses', 'terhadap', 'sistem', 'elektronik', 'access', 'blocking', 'menteri', 'melakukan', 'normalisasi', 'berdasarkan', 'pengajuan', 'rekomendasi', 'oleh', 'kementerian', 'atau', 'lembaga', 'atas', 'dasar', 'layanan', 'pse', 'lingkup', 'privat', 'yang', 'telah', 'memenuhi', 'ketentuan', 'peraturan', 'perundangundangan']

4. Stemming

In the Stemming process, datasets containing affixes to each word will be converted to base words [14]. In this process, each word fraction of the tokenizing process will be normalized into a root word. In this step, the conjunctions or conjunctions will be removed. The output of stemming is described in table 4.

Table 4. Application of Stemming in Datasets.

Before	After
['dalam', 'hal', 'sanksi', 'administratif', 'yang', 'diberikan', 'kepada', 'pse', 'lingkup', 'privat', 'sebagaimana', 'dimaksud', 'pada', 'ayat', 'adalah', 'pemutusan', 'akses', 'terhadap', 'sistem', 'elektronik', 'access', 'blocking', 'menteri', 'melakukan', 'normalisasi', 'berdasarkan', 'pengajuan', 'rekomendasi', 'oleh', 'kementerian', 'atau', 'lembaga', 'atas', 'dasar', 'layanan', 'pse', 'lingkup', 'privat', 'yang', 'telah', 'memenuhi', 'ketentuan', 'peraturan', 'perundangundangan']	['sanksi', 'administratif', 'pse', 'lingkup', 'privat', 'ayat', 'putus', 'akses', 'sistem', 'elektronik', 'access', 'blocking', 'menteri', 'normalisasi', 'dasar', 'aju', 'rekomendasi', 'menteri', 'lembaga', 'dasar', 'layan', 'pse', 'lingkup', 'privat', 'penuh', 'tentu', 'atur', 'perundangundangan']

Sentiment Labeling

Sentiment labeling uses a lexicon-based approach to datasets preprocessed by Text Preprocessing using the InSet Lexicon [15]. Assessment is done by calculating the value of each word by matching the existing words from each sentence in the word dictionary dataset that has been given a weight. After each word gets a weight value, a calculation is performed to get each sentence's total value (polarity score) to be given a sentiment label. In general, the sentiment label is stated as "Negative" if the polarity score obtained is below 0, the "Positive" label is given if the polarity score obtained is above 0, and the "Neutral" label is given if the polarity score has a value of 0. The output of the sentiment prediction labeling process with the InSet Lexicon is described in table 5.

Table 5. Sentiment Analysis Results Based on Compound Score

Text	Polarity Score	Sentiment Prediction
selenggara jasa akses internet internet service provider singkat isp selenggara jasa multimedia selenggara jasa layan akses internet hubung jaring internet publik	28	Positive
menteri kena sanksi administratif pse lingkup privat daftar pasal pasal tanda daftar lapor ubah informasi daftar pasal informasi daftar pasal ayat pasal ayat pasal ayat	-36	Negative
selenggara sistem elektronik orang selenggara negara badan usaha masyarakat sedia kelola operasi sistem elektronik sendirisendiri bersamasama guna sistem elektronik perlu perlu	0	Neutral

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 8. NO. 1 AGUSTUS 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4091

Sentiment analysis was performed on a dataset consisting of 8 article lines. The sentiment prediction graph for the dataset is displayed in Figure 2, which shows the sentiment distribution for the articles in the dataset based on the InSet Lexicon dictionary. This dictionary commonly categorizes words in a text into positive, negative, or neutral sentiment categories. The graph indicates that the majority of the articles in the dataset had a negative sentiment, suggesting that the articles tended to convey messages or information that were less joyful or poignant. Articles with positive and neutral sentiments were present in smaller proportions. The results showed that the majority of the articles, 80.34%, had a predicted negative sentiment, 17.42% had a positive sentiment, and 2.25% had a neutral sentiment.

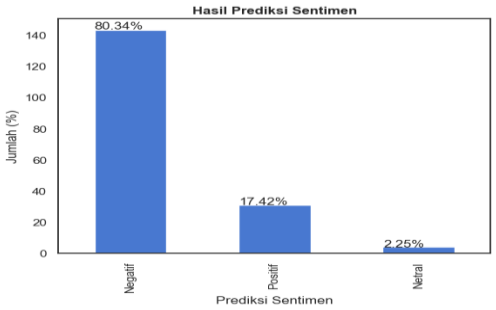

Figure 2. Sentiment Predicted Distribution of Datasets with InSet Lexicon

The sentiment analysis results implicitly indicate that articles 7 and 8 of Ministerial Regulation Number 5 of 2020 potentially create legal uncertainty for the public in accessing content, which is consistent with the research findings showing negative results. The following results from sentiment prediction analysis on articles 7 and 8 of Ministerial Regulation Number 5 of 2020 on Communication and Information Technology, as described in Table 6.

Table 6. Results of Sentiment Prediction in Articles 7 and 8

Articles	Section	Text Preprocessing	Score
7	1	pse lingkup privat daftar ayat huruf menteri sanksi administratif putus akses sistem elektronik access blocking	-7
7	2	pse lingkup privat tanda daftar lapor ubah informasi daftar ayat huruf informasi daftar ayat huruf menteri sanksi administratif tegur tulis surat elektronik electronic mail media elektronik henti pse lingkup privat indah tegur tulis ayat huruf putus akses sistem elektronik access blocking cabut tanda daftar selenggara sistem elektronik pse lingkup privat konfirmasi jangka tujuh henti ayat huruf	-30
7	3	pse lingkup privat penuh tentu daftar pasal pasal menteri normalisasi sistem elektronik putus akses access blocking ayat	-18
7	4	pse lingkup privat baru informasi daftar menteri normalisasi sistem elektronik henti ayat huruf	-14
7	5	pse lingkup privat daftar ulang informasi daftar menteri normalisasi sistem elektronik putus akses sistem elektronik cabut tanda daftar selenggara sistem elektronik ayat huruf	-15
7	6	menteri kena sanksi administratif pse lingkup privat dasar mohon menteri lembaga dasar langgar atur perundangundangan bidang menteri lembaga milik wenang sesuai tentu atur perundangundangan	-3
8	1	sanksi administratif pse lingkup privat ayat putus akses sistem elektronik access blocking menteri normalisasi dasar aju rekomendasi menteri lembaga dasar layan pse lingkup privat penuh tentu atur perundangundangan	-16
8	2	pse lingkup privat tanggung selenggara sistem elektronik kelola informasi elektronik dokumen elektronik sistem elektronik andal aman tanggung	-13

The sentiment analysis on Articles 7 and 8 shows adverse outcomes and implicitly indicates the potential for legal uncertainty for the public in accessing content, consistent with the research findings. These results are supported by validation from two legal experts who deemed the analysis results valid and by prevailing conditions. The legal expert confirmed that it was conducted accurately and in line with legal principles. The experts

CC BY NC

considered the analysis results reliable and accurate.

The results of the sentiment prediction are visualized in graphical form. WordCloud is a method for visualizing words based on the number of occurrences or their frequency in a document [16]. The more often the word is used, the greater the visualization of the word. The following is a word visualization based on the frequency of word use in the content of PERMENKOMINFO No. 5 of 2020, which is shown in figure 3.

Figure 3. WordCloud for PERMENKOMINFO Content No.5 of 2020

The results of the word visualization applied in the WordCloud method show that the words "akses" and "sistem" are the most frequently used in the content of PERMENKOMINFO No. 5 of 2020. This indicates how vital both words are in the context of the regulation and suggests that access and system are the main focus of the regulation. This visualization also provides valuable information to understand the context and main focus of the content of the regulation. The frequency of word occurrences is also explained in Figure 4 based on the 15 words with the most usage in the dataset for the article on PERMENKOMINFO No.5 of 2020.

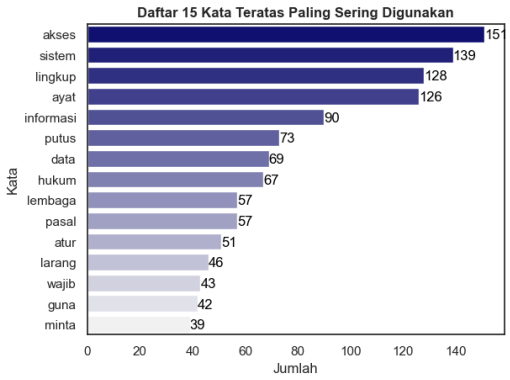

Figure 4. List of Popular Words that Often Appear

After the sentiment analysis has been conducted and approved by legal experts, the next step is to assign weights to the analyzed words that have been given sentiment values using the TF-IDF method. Term Frequency -Inverse Document Frequency is a statistical analysis technique that analyses the frequency of occurrence of words and is combined with inverse [17]. The purpose of weighting words using the TF-IDF method on the sentiment analysis results of the content of Ministerial Regulation No. 5 of 2020 is to provide different weights to each word in the document based on its meaning in the text. This weighting helps identify the most critical words in the document and the sentiment polarity. The results of the TF-IDF calculations on the dataset are shown in Figure 5.

	TF-IDF
access	3.142863
acu	5.494239
adil	3.884801
administratif	3.548328
adu	4.801091
...	...
web	5.494239
website	4.801091
wenang	3.622436
wilayah	3.990161
yurisdiksi	5.494239

439 rows × 1 columns

Figure 5. TF-IDF Calculation Results on the PERMENKOMINFO Content Dataset No. 5 of 2020

After performing TF-IDF weighting, the next step in sentiment analysis research on the content of Ministerial Regulation No. 5 of 2020 is to perform a data split. Data split is done to divide the data into two parts: training data and testing data. The training data is used to train the data classification model, while the testing data is used to test the performance of the trained model. After the data is split, the next step is to use data classification methods, namely Naive Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors. These three methods are used to predict the sentiment of the given text, whether the previously predicted sentiment results are tested to see if they are positive, negative, or neutral. The NBC method calculates the probability of each sentiment class based on the occurrence of words in the text, while the SVM method uses optimization theory-based learning algorithms and linear functions in multidimensional features. The KNN method works by grouping new data based on the distance between the new data and some of the nearest data.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 8. NO. 1 AGUSTUS 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4091

In this study, all three data classification methods were used to test the sentiment of the content analysis of Ministerial Regulation No. 5 of 2020 and to test the performance of the trained model using testing data. The following is the model testing results for the sentiment analysis on the content of Ministerial Regulation No. 5 of 2020.

2.4. Support Vector Machine

Test ratio of 80:20 using the SVM algorithm. The result of the test accuracy is 83%. The result of the precision value is 82%. The results of the recall value are 100%, and the f1-score value is 90%. Complete results are shown in Figure 6.

	precision	recall	f1-score	support
Negatif	0.82	1.00	0.90	28
Netral	0.00	0.00	0.00	2
Positif	1.00	0.33	0.50	6
accuracy			0.83	36
macro avg	0.61	0.44	0.47	36
weighted avg	0.81	0.83	0.79	36

Figure 6. Testing Result of Support Vector Machine

2.5. Naïve Bayes Classifier

Tests carried out with the NBC algorithm with a test ratio of 80:20 obtained 78% accuracy, 78% precision, 100% recall, and 88% f1-score, explained in Figure 7.

	precision	recall	f1-score	support
Negatif	0.78	1.00	0.88	28
Netral	0.00	0.00	0.00	2
Positif	0.00	0.00	0.00	6
accuracy			0.78	36
macro avg	0.26	0.33	0.29	36
weighted avg	0.60	0.78	0.68	36

Figure 7. Testing Result of Naive Bayes Classifier

2.6. K-Nearest Neighbors

The test results with the KNN model, with a test ratio of 80:20. Obtained an accuracy value of 83%, a precision value of 87%, a recall value of 93%, and an f1-score value of 90%, which can be seen in Figure 8.

	precision	recall	f1-score	support
Negatif	0.87	0.93	0.90	28
Netral	0.00	0.00	0.00	2
Positif	0.67	0.67	0.67	6
accuracy			0.83	36
macro avg	0.51	0.53	0.52	36
weighted avg	0.79	0.83	0.81	36

Figure 8. Testing Result of K-Nearest Neighbors

RESULTS AND DISCUSSION

The results of the data classification testing showed that the use of the Naive Bayes method provided a testing accuracy of 78%, while the SVM and KNN methods provided a testing accuracy of 83%. This testing was conducted on 178 articles, focusing on articles 7 and 8. The results of the comparison of method testing are described in Table 7.

Table 7. Test Results for Each Classification Method

Method / Rated	Support Vector Machine	Naïve Bayes	K-Nearest Neighbors
accuracy	83%	78%	83%
precision	82%	78%	87%
recall	100%	100%	93%
f1-score	90%	88%	90%

This indicates that using data classification methods in sentiment analysis can provide accurate results in predicting the sentiment towards the content of PERMENKOMINFO No.5 Tahun 2020, especially in articles 7 and 8 which are the main focus of the research. With good accuracy results, this study can contribute significantly to understanding public sentiment towards the policy. These results indicate that the research was conducted correctly and validly. The fact that the results were also supported by validation from two legal experts strengthens the statement that the sentiment analysis conducted was by legal standards and produced correct results. Both data classification and legal expert validation results can be concluded that the policy in PERMENKOMINFO No.5 Tahun 2020 related to legal uncertainty in accessing content is considered harmful by the public. Therefore, there is a need for improvement and better policies to ensure the interests of the public in accessing online information.

CONCLUSION

The conclusion of this study is that the data classification method in sentiment analysis can provide accurate results in predicting the sentiment toward the content of PERMENKOMINFO No.5 of 2020. The study also shows the need for improvement and better policies to ensure the interests of the public in accessing online information. The negative sentiment result of 80.34% obtained in the study contributes to being an input for evaluation. This study provides valuable insights into the public's sentiment towards the policy in PERMENKOMINFO No.5 of 2020 and its impact on their ability to access the content. The study contributes to the understanding

CC BY NC

of legal uncertainty in accessing content and the need for better policy to ensure the interests of the public.

REFERENCE

[1] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, Jan. 2023.

[2] M. Pilliang, H. Akbar, and G. Firmansyah, "Sentiment Analysis for Super Applications in Indonesia: A Case Study of Gov2Go App," *Proc. Int. Conf. Electr. Eng. Informatics*, vol. 2022-October, pp. 80–85, 2022.

[3] I. G. N. D. Adnyana, F. Adams, A. W. Oktavia, E. Ermatita, and S. Sarika, "Analisis Sentimen Terhadap Undang-Undang Cipta Kerja Menggunakan Metode Naïve Bayes," in *SENAMIKA*, 2021, vol. 2, no. 2, pp. 120–129.

[4] P. Sejati, M. Munawar, M. Pilliang, and H. Akbar, "Studi Komparasi Naive Bayes, K-Nearest Neighbor, dan Random Forest untuk Prediksi Calon Mahasiswa yang Diterima atau Mundur," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 7, pp. 1341–1348, Dec. 2022.

[5] Y. V. Wijaya, A. Erfina, and C. Warman, "Analisis Sentimen Seputar UU ITE Menggunakan Algoritma Support Vector Machine," *Progresif J. Ilm. Komput.*, vol. 17, no. 2, pp. 1–14, Aug. 2021.

[6] E. N. Hamdana, M. Balya, and I. Alfahmi, "Pengembangan Sistem Analisis Sentimen Berbasis Java Pada Data Twitter Terhadap Omnibus Law Menggunakan Algoritma Naïve Bayes dan K-Nearst Neighbor (K-NN)," *J. Inform. Polinema*, vol. 7, no. 2, pp. 79–84, Feb. 2021.

[7] A. Syaifuddin and M. Muslimin, "Analisis Sentimen Pada Sosial Media Tentang Implementasi Kebijakan PSE Kominfo Menggunakan Algoritme Lexicon Based," in *Seminar Nasional Fakultas Teknik*, 2022, vol. 1, no. 1, pp. 7–14.

[8] A. Rahman, W. Wiranto, and A. Doewes, "Online News Classification Using Multinomial Naive Bayes," *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 32–38, Aug. 2017.

[9] R. Rismayani, H. SY, T. Darwansyah, and I. Mansyur, "Implementasi Algoritma Text Mining dan Cosine Similarity untuk Desain Sistem Aspirasi Publik Berbasis Mobile," *Komputika J. Sist. Komput.*, vol. 11, no. 2, pp. 169–176, Aug. 2022.

[10] D. S. Suparno and M. Rosyda, "Penggunaan Text Modeling Untuk Identifikasi Kesalahan Penulisan Kata Pada Teks Pidato Bupati Banggai Sulawesi Tengah," *J. Media Inform. Budidarma*, vol. 5, no. 3, pp. 779–789, Jul. 2021.

[11] Y. Afrillia, L. Rosnita, and D. Siska, "Analisis Sentimen Pengguna Twitter Terhadap Isu Kesetaraan Gender Dalam Penerapan Permendikbudristek Nomor 30 Tahun 2021 Menggunakan Textblob Analysis Of Twitter User Sentiment Towards To Issue Of Gender Equality In The Implementation Of Permendikbudristek Number 30 Of 2021 Using Textblob," *J. INFORMATICS Comput. Sci.*, vol. 8, no. 2, pp. 93–98, 2022.

[12] P. Aditiya, U. Enri, and I. Maulana, "Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, pp. 1020–1028, Aug. 2022.

[13] S. Mujahidin, B. Prasetio, and M. C. C. Utomo, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian Naïve Bayes," *Voteteknika (Vocational Tek. Elektron. dan Inform.*, vol. 10, no. 3, pp. 17–24, Sep. 2022.

[14] O. I. Gifari, M. Adha, I. Rifky Hendrawan, and F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, Mar. 2022.

[15] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-January, pp. 391–394, Feb. 2018.

[16] M. G. Pradana, "Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining," *J. Ilm. Inform.*, vol. 8, no. 01, pp. 38–43, Mar. 2020.

[17] A. M. Priyatno and L. Ningsih, "TF-IDF Weighting to Detect Spammer Accounts on Twitter based on Tweets and Retweet Representation of Tweets," *Sist. J. Sist. Inf.*, vol. 11, no. 3, pp. 614–622, Sep. 2022.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri