

Research Article

## Social Media Sentiment Analysis of Instagram Use by Early Childhood Education Information System Development Based on Naïve Bayes

Yuma Akbar<sup>1\*</sup>, Sugiyono<sup>2</sup>, Dedi Gunawan<sup>3</sup>, Salsabila Putri Wibowo<sup>4</sup>

<sup>1</sup>Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia;  
email: [yumekhan@stikomcki.ac.id](mailto:yumekhan@stikomcki.ac.id)

<sup>2</sup>Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia;  
email: [sugiyono@stikomcki.ac.id](mailto:sugiyono@stikomcki.ac.id)

<sup>3</sup>Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia;  
email: [dedigunawan@stikomcki.ac.id](mailto:dedigunawan@stikomcki.ac.id)

<sup>4</sup>Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia

\*Corresponding Author: [yumekhan@stikomcki.ac.id](mailto:yumekhan@stikomcki.ac.id)

**Abstract:** This study employs the Naïve Bayes method to analyze social media sentiment regarding the use of Instagram by early childhood users. The primary objective of this research is to understand public perceptions of the positive and negative impacts of Instagram usage among young children, particularly in relation to their social, psychological, and digital behavioral development. Sentiment analysis is carried out using data from various social media platforms, which are then classified into positive, negative, and neutral opinions. The classification results form the basis for developing an integrated educational information system designed to provide guidance for parents, educators, and children in using Instagram safely, healthily, and responsibly. The system also emphasizes the importance of age-appropriate content education, privacy settings, and strategies to minimize the risks of exposure to inappropriate content and the negative effects of excessive usage. This research is expected to support the creation of a more positive, safe, and beneficial digital environment for early childhood users while also serving as a reference in formulating effective policies in the social media era.

**Keywords:** Early Childhood; Educational; Information Systems; Instagram Social Media; Naïve Bayes Method.

Received: 16 December 2025

Revised: 10 January 2025

Accepted: 12 February 2026

Published: 20 February 2026

Curr. Ver.: 20 February 2026



Copyright: © 2025 by the authors.

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license

(<https://creativecommons.org/licenses/by-sa/4.0/>)

### 1. Introduction

The development of digital technology has brought significant changes to human social interactions in the modern era. Social media, as a form of communication technology, has experienced rapid growth and become an inseparable part of daily life. Instagram, as one of the social media platforms emphasizing visual content, has achieved remarkable popularity, with more than 2 billion monthly active users worldwide as of 2024 (Aditya et al., 2022).

The use of Instagram by young children has generated various concerns and opinions within society. Some parties emphasize the potential negative impacts, such as exposure to inappropriate content, the risk of cyberbullying, and disruptions to social development. On the other hand, there is a view that, with proper supervision, social media can also provide educational and developmental benefits for children (Ula & Fachrurrazi, 2023).

Sentiment analysis is a branch of computer science that aims to identify, extract, and quantify individuals' opinions and attitudes toward specific entities, events, or topics. In the context of social media, sentiment analysis can be utilized to map public opinions regarding

the use of Instagram by young children. The results of such analyses can provide valuable insights for developing policies, educational programs, and information systems that assist parents and educators in guiding children toward healthy social media usage (Pratidina & Mitha, 2023).

The Naïve Bayes classification method was selected in this study due to its advantages in text sentiment analysis. This algorithm is well known for its computational efficiency, ability to work with relatively small datasets, and effectiveness in classifying Indonesian-language text. Through a probabilistic approach, the Naïve Bayes method can classify public opinions into positive, negative, or neutral sentiments based on the linguistic features present in the dataset (Susanti et al., 2022).

Based on this background, this study aims to develop an educational information system based on sentiment analysis using the Naïve Bayes method. The system is expected to provide comprehensive information regarding public perceptions of Instagram usage among young children while also serving as an educational platform for parents and educators in guiding children to use social media in a healthy and responsible manner.

This research focuses on the development of an educational information system based on sentiment analysis. The study employs sentiment analysis methods to identify public opinions related to Instagram use among young children. The sentiment analysis utilizes the Naïve Bayes algorithm (Ula & Fachrurrazi, 2023), which was chosen for its computational efficiency and effectiveness in processing Indonesian-language text. The results of the sentiment analysis are then integrated into an educational information system designed to provide comprehensive information and educational resources for parents and educators regarding healthy and responsible social media usage. In other words, this study adopts a quantitative approach (sentiment analysis) to generate data that are subsequently interpreted and presented through a qualitative educational information system (Zamzami et al., 2024).

Several previous studies serve as references for the development of this system. First, research on the use of data mining for business strategy in the culinary sector has continued to grow, particularly through the integration of RFM analysis and the Apriori algorithm. RFM is used for customer segmentation based on transaction behavior, while Apriori identifies product combinations frequently purchased together. Previous studies have demonstrated the effectiveness of these methods, such as Merliani et al. (2022) in menu recommendation systems, and Susilowati et al. (2022) and Widiyanto and Witanti (2021), who combined them with clustering techniques. In a study conducted at Bonjour Café and Resto using transaction data from December 2024, a strong association was identified among “Grilled Tilapia,” “Vit 550 ml,” and “French Fries” (support = 28%, confidence = 70%), along with the segmentation of five customer categories. The primary contribution of this study lies in the combined application of RFM and Apriori with a monthly analytical focus in the culinary industry, an area that has received limited attention in existing literature, as well as the application of the CRISP-DM approach to ensure systematic and valid analysis (Marzuki & Sibarani, 2024).

Second, a study analyzed public sentiment on Twitter regarding the Indonesian government’s closure of TikTok Shop. Using the Naïve Bayes algorithm and labeled data derived from the VADER lexicon, the findings indicated that most netizens responded positively to the policy. The developed model achieved an accuracy of 85%, with a precision of 99.33% and a recall of 77.20%. This study contributes to the understanding of digital public opinion and is relevant to evidence-based policy-making using social media data (Sanjaya et al., 2024).

Third, another study examined public opinion toward the LGBT community on Twitter using the Naïve Bayes algorithm. Of the 1,579 collected tweets, 1,502 were classified as negative and 77 as positive. The model achieved accuracies of 95.24% and 87.58% through cross-validation. These findings indicate the dominance of negative opinions toward LGBT individuals on Twitter while also demonstrating the effectiveness of the Naïve Bayes method in social opinion classification (Anjani et al., 2022).

This study offers an original contribution by developing an educational information system based on sentiment analysis using the Naïve Bayes method to map public opinions regarding Instagram usage among young children. Its uniqueness lies in the application of this method within this specific context, which has received limited attention in previous research (Devi & Fadli, 2023).

Although sentiment analysis has been widely applied in various contexts, its application to mapping public opinion about children's use of social media, particularly Instagram, and providing an educational information system based on the analysis results is relatively novel. Previous studies may have examined sentiments related to social media or children's internet usage; however, few have integrated sentiment analysis with the development of an educational information system specifically targeted at parents and educators (Zuraiyah et al., 2023).

The resulting information system is expected to provide comprehensive and easily understandable information regarding public perceptions of the positive and negative impacts of Instagram usage among children. This information can assist parents and educators in making more informed decisions regarding the supervision and guidance of underage children in using social media. Consequently, this study contributes to the enhancement of digital literacy and child-rearing practices in the digital era (Laila & Yamin, 2023).

## 2. Literature Review

### Early Childhood Use of Social Media

Early childhood is a period characterized by rapid development in physical, cognitive, and socio-emotional aspects. The use of social media by young children is a complex phenomenon involving children's access to and interaction with digital platforms such as YouTube, Instagram, TikTok, and others. This usage may vary, ranging from simply watching videos to actively engaging with content and interacting with other users. It is important to note that such access and usage often occur under varying levels of parental supervision (Adelia et al., 2023).

### Sentiment Analysis

Sentiment analysis is a method within the fields of Natural Language Processing (NLP), data mining, and artificial intelligence that aims to identify, extract, and classify opinions, emotions, or attitudes toward a particular entity, topic, or event based on textual data. In other words, sentiment analysis is used to determine whether a statement conveys positive, negative, or neutral sentiment. This technique is commonly applied to data obtained from social media platforms, product reviews, comments, and online news sources.

The sentiment analysis process generally involves several stages, including data collection, text cleaning (text preprocessing), feature extraction, and classification using machine learning algorithms, one of which is Naïve Bayes. In this study, sentiment analysis is employed to map public opinions regarding the use of Instagram by young children. The results not only provide an overview of public perceptions but can also serve as a foundation for policy formulation, digital education strategies, and the development of information systems that assist parents and educators in guiding children toward healthy and responsible social media usage.

### Instagram

Instagram is a photo- and video-based social networking service that enables users to share multimedia content, interact with friends and followers, and follow accounts that match their interests. Users can upload photos and videos, edit them using various filters and effects, and add captions and hashtags to increase the visibility of their content. Instagram is available across multiple devices, including smartphones (iOS and Android), computers, and web browsers.

The platform is widely used by both individuals and businesses to share personal moments, promote brands, and build online communities. However, it is important to use Instagram responsibly and to be aware of its potential negative impacts, such as addiction and cyberbullying (Setiawan & Widyaningrum, 2024). In addition to serving as a source of entertainment, Instagram also functions as a medium for communication, promotion, and learning. Features such as Stories, Reels, Live Streaming, and Direct Messages provide interactive experiences that encourage active user participation.

### Naïve Bayes

Naïve Bayes is a classification method in machine learning based on the application of Bayes' Theorem with the "naïve" assumption that each feature (attribute) in a dataset contributes independently to the probability of a particular class. Although this assumption of independence is rarely fully satisfied in real-world situations, Naïve Bayes often produces strong performance across a wide range of practical applications (Zusrotun et al., 2022).

$$P(X|H) = \frac{P(X|H)P(H)}{P(H)}$$

X : Data with an unknown class.

H : The hypothesis that data X belongs to a specific class.

P(H|X) : The probability of hypothesis H given the condition X.

P(H) : The probability of hypothesis H.

P(X|H) : The probability of observing X given hypothesis H.

P(X) : The probability of observing X.

The Naïve Bayes algorithm works by calculating the probability that a piece of data belongs to a particular class based on the distribution of its features. Its probabilistic nature makes this method simple, fast, and efficient for application to both large and small datasets. One of the primary advantages of Naïve Bayes is its ability to address text classification problems, such as sentiment analysis, email spam filtering, and product recommendation systems. Another advantage of the algorithm is its ability to perform well even when dealing with data imbalance, while its classification results are relatively easy to interpret. However, this method also has limitations, particularly because its assumption of feature independence may not always hold in real-world data, which can affect prediction accuracy. Therefore, effective data preprocessing is essential to optimize classification performance.

### Data

In computer science and machine learning, a dataset is a collection of data typically organized in a tabular format, where each row represents an entity or record, and each column represents a feature or attribute of that entity. Datasets are used for various purposes, including data analysis, machine learning model training, and algorithm testing (Pratidina & Mitha, 2023).

### RapidMiner

RapidMiner is a software platform designed for data analysis, machine learning, and data mining. It provides comprehensive tools for all stages of the data analysis process, including data collection, preprocessing, modeling, evaluation, and predictive model deployment. RapidMiner is widely recognized for its ease of use, intuitive interface, and support for various machine learning algorithms (Harahap & Kurniawan, 2024).

As one of the most widely used open-source data mining tools, RapidMiner supports multiple data formats and can be integrated with programming languages such as Python and R. This flexibility makes it suitable for both academic research and industrial applications. Its drag-and-drop functionality enables users, including beginners with limited programming experience, to analyze complex datasets efficiently.

In the context of this study, RapidMiner is utilized to support the sentiment analysis process using the Naïve Bayes algorithm, covering stages such as data cleansing, tokenization, and sentiment classification into positive, negative, and neutral categories. With the assistance of RapidMiner, the analysis process can be conducted more quickly, systematically, and with improved visualization, thereby enhancing the transparency and reliability of the research results.

Another advantage of RapidMiner is the availability of various extensions that support big data analytics, integration with relational databases, and data visualization techniques. However, one of its limitations is the restricted functionality available in the free version, particularly for processing large-scale datasets. Nevertheless, RapidMiner remains one of the preferred tools among researchers and data practitioners for effective data processing and analysis.

### Social Media

Media refers to a means or tool used to convey messages, information, or ideas from one party to another, facilitating the communication process. The term "media" originates from the Latin word meaning "intermediary" or "connector." In the context of communication, media serve as a bridge that enables messages to be delivered, received, and understood effectively.

Media can be categorized into several types, including print media (such as newspapers, magazines, and books), electronic media (such as television and radio), and digital media (such as the internet, social media platforms, and technology-based applications). Technological advancements have accelerated the transition from traditional media to digital media, which are now more interactive, faster, and capable of reaching users on a global scale.

Social media, in particular, is viewed as a modern communication medium that enables users to share information in the form of text, images, and videos while facilitating interaction and collaboration among individuals and communities.

### System Development

System development is a structured process aimed at designing, building, implementing, and maintaining a system to meet the needs of users and organizations. In this context, a system may refer to an information system, software application, or a combination of technological components, people, and procedures working together to achieve specific objectives.

The system development process generally consists of several stages, including requirements analysis, system design, implementation, testing, and maintenance. Common methodologies used in system development include the System Development Life Cycle (SDLC), Agile Development, and Rapid Application Development (RAD).

The primary objective of system development is to improve effectiveness, efficiency, and service quality through the appropriate use of technology. In this study, system development is carried out to design an educational information system based on sentiment analysis. The system is intended to provide guidance and educational resources for parents and educators in assisting young children to use Instagram and other social media platforms in a healthy, safe, and responsible manner.

## 3. Method Study

### Research Data

The research data for this study were specifically collected from Instagram discussions related to the use of Instagram by young children. Instagram was selected as the primary data source because it is one of the most popular social media platforms and is widely used by parents to share content about their children. The dataset consisted of comments collected from posts featuring early childhood-related content, resulting in a total of 435 comments (Rosidi & Setiawan, 2024).

### Method Implementation

At this stage of the research, initial data collection was conducted, followed by data retrieval. The collected data then underwent a data cleansing process, followed by attribute selection. Subsequently, the CRISP-DM methodology was applied, consisting of six stages that guided the overall research process (Adelia et al., 2023).

The implementation of the method in this study was carried out through several systematic stages, beginning with data collection and ending with sentiment analysis using the Naïve Bayes algorithm. The stages are described as follows.

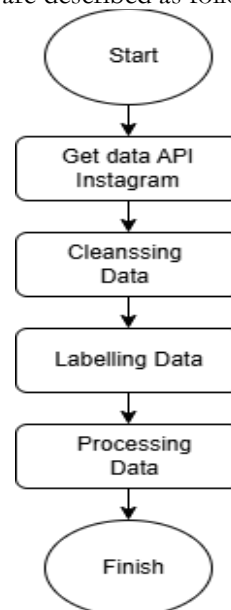


Figure 1. Research Framework.

Based on the figure above, this study consists of four main processes:

a) GET Data Instagram

The object of this study is Instagram users' opinions regarding young children. The data were collected from Instagram content featuring young children.

b) Cleasing Data

The collected hashtags and textual data were processed through several stages, including cleaning, tokenization, case transformation, stopword removal, and filtering.

c) Labeling Data

After completing the data cleansing process, the data were manually labeled. Manual labeling was applied to all hashtags and textual content to ensure accurate sentiment classification.

d) Processing Data

This study employed a confusion matrix to evaluate the performance of the classification model. The testing process was conducted on the data generated during the modeling stage using the Naïve Bayes algorithm to assess its classification accuracy.

### **Testing Design**

#### ***Business Understanding***

This stage focuses on understanding the research objectives related to the use of Instagram by young children. The primary problem identified is the need to classify Instagram content as either safe or unsafe for children aged 3–6 years. The business objectives include developing an automated filtering system capable of protecting children from inappropriate content, assisting parents in monitoring their children's digital activities, and creating a safer social media environment. The key stakeholders in this project include parents, early childhood educators, parental control application developers, and social media platforms seeking to enhance the safety of their younger users (Anjani et al., 2022).

#### ***Data Understanding***

At this stage, the Instagram data intended for training the Naïve Bayes model were explored and analyzed. The collected data included captions, hashtags, comments, image metadata, and user profile information from Instagram posts. The data exhibited various characteristics, including the use of Indonesian and English languages, emojis, internet slang, and different types of multimedia content (Wahyuni, 2022).

#### ***Data Preparation***

The process began with text cleaning, including the removal of URLs, excessive hashtags, and irrelevant special characters. Text normalization was then performed through lowercase conversion and the removal of Indonesian and English stopwords.

Data labeling was conducted manually to ensure the accuracy of the "safe" and "unsafe" categories according to content standards appropriate for early childhood audiences (Setiawan & Widyaningrum, 2024).

#### ***Modeling***

The modeling stage employed a classification approach using the Naïve Bayes algorithm. The classification model was developed to categorize content related to young children. RapidMiner was utilized as the primary modeling tool due to its capability to support machine learning and text mining processes efficiently (Mubarok & Adjani, 2025).

#### ***Evaluation (Evaluasi)***

At this stage, the accuracy and effectiveness of the developed model were analyzed and measured. The evaluation aimed to determine whether the model was appropriate and suitable for the research case and whether it aligned with the initial research objectives. Based on the evaluation results, a decision was made regarding whether the process could proceed to the next stage or whether revisions and reprocessing were required due to inconsistencies with the original research plan (Heristian et al., 2025).

#### ***Deployment***

The final stage involved disseminating the results of the study in the form of reports and presentations. The knowledge generated from the modeling and evaluation stages of the data mining process was documented and presented to facilitate understanding and future implementation of the research findings.

## 4. Results and Discussion

### Research Instruments

This study requires several instruments to support the research process. The research instruments used consist of both software and hardware components. These tools are utilized to facilitate data collection, data processing, sentiment analysis, system development, and the overall implementation of the research.

#### Software

The following software tools, their versions, and their functions are presented in Table 2.

**Table 2.** Software Specifications.

No.	Software Type	Version	Function
1	Microsoft Excel	2010	Used for data filtering and data storage.
2	RapidMiner	10.2	Used for data processing and analysis.
3	Mendeley	1.19.8	Facilitates the scientific writing process, including reference management, citation generation, and automatic bibliography creation.

#### Hardware

The hardware tools, their types, and specifications are presented in Table 3.

**Table 3.** Hardware Specifications.

No	Hardware Type	Specification
1	<i>Processor</i>	Inter(R) Celeron(R)N420 CPU 1.010Ghz (2CPUs)
2	RAM	4GB
3	<i>System Type</i>	Windows
4	<i>Operating System</i>	Windows 11 Home

### Implementation and Testing

The implementation and testing of the Naïve Bayes method were conducted using data collected from comments on early childhood-related content. The dataset consisted of 435 comments categorized into positive, negative, and neutral sentiments.

#### a) Data Collection

To collect data using Excel, the number of data analyzed in this study was 435 consisting of positive, negative and neutral data as shown in Figure 2.

	A	B	C	D	E
1	nomor	Bahasa Baku			
2	1	menggunakan jaket kita ke rumah rumah sakit jiwa	negatif		
3	2	kerusakannya itu loh sangat tidak ada	negatif		
4	3	biarin saja nanti juga besar malu sendiri	negatif		
5	4	bertato saja bangga palingan juga dari permen karet	negatif		
6	5	kamu kalau mau garuk idung bilang tidak usah di rekam video	negatif		
7	6	sangat tidak nyambung banget	negatif		
8	7	wah sangat keren dan sangat cantik	positif		
9	8	auranya itu loh sangat tidak ada	negatif		
10	9	seperti orang gila bikin kayak seperti itu	negatif		
11	10	sungguh tidak menarik	negatif		
12	11	mengapa dengan dia	negatif		
13	12	untuk apa dia berperilaku seperti itu	negatif		
14	13	saya sangat merinding asli	negatif		
15	14	sungguh menjijikan membuat seperti itu	negatif		
16	15	suara dan nada terpisah	negatif		
17	16	yang tertawa bukan hanya langit	positif		
18	17	nada sesuai kepercayaan masing-masing	positif		
19	18	besok akan menjadi meme ini	negatif		
20	19	saya tidak mau kembali kesini lagi	negatif		
21	20	apa malam malam bikin heboh	negatif		
22	21	ini ceritanya membuat apa	negatif		
23	22	aduh sangat takut kepada adik ini	negatif		

**Figure 2.** Amount of Data Obtained.

#### b) Cleansing Data

At this stage, the researcher will carry out the Cleansing process using Rapidminer before labeling the sentiment, so that the dataset in the latest Excel file is free from data duplication and unnecessary marks as depicted in Figure 3.

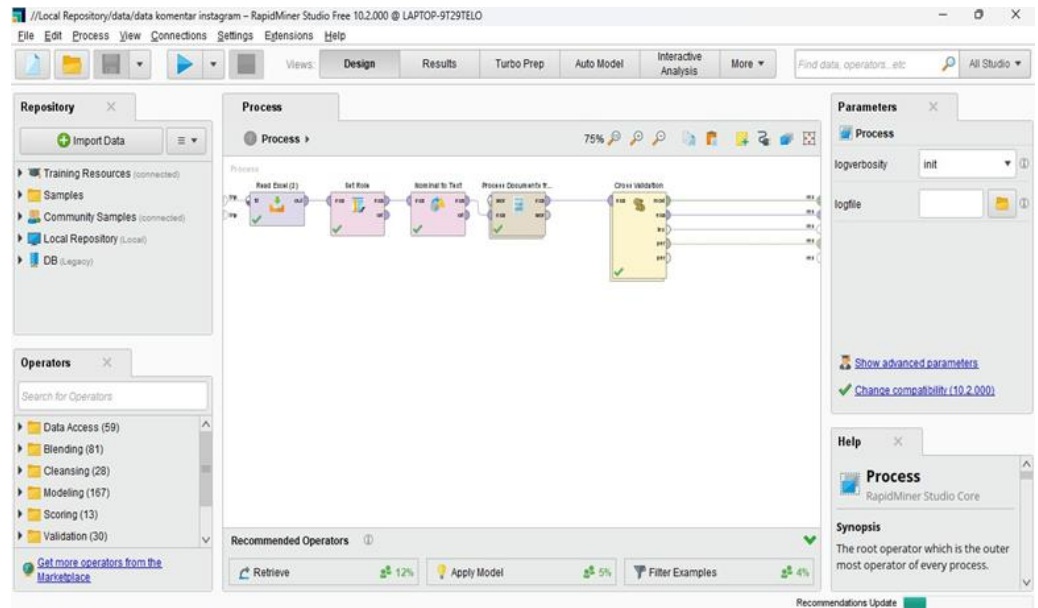


Figure 3. Cleansing Process Model.

After completing the Cleansing process using RapidMiner, the data obtained is as shown in Figure 4.

Open in Turbo Prep Auto Model Filter (434 / 434 examples): all

Row No.	hasil	text	abang	abi	abisan	adakah	adalah	adik
1	negatif	menggunaka...	0	0	0	0	0	0
2	negatif	kerusakanny...	0	0	0	0	0	0
3	negatif	biarin saja na...	0	0	0	0	0	0
4	negatif	bertato saja b...	0	0	0	0	0	0
5	negatif	kamu kalau g...	0	0	0	0	0	0
6	negatif	sangat tidak ...	0	0	0	0	0	0
7	positif	sangat keren ...	0	0	0	0	0	0
8	negatif	auranya sang...	0	0	0	0	0	0
9	negatif	seperti orang...	0	0	0	0	0	0
10	negatif	sungguh tida...	0	0	0	0	0	0
11	negatif	mengapa de...	0	0	0	0	0	0
12	negatif	unbuh berpadi...	0	0	0	0	0	0

Figure 4. Data Cleansing Results.

c) Labeling

The next step is to classify or determine the current sentiment. The following data has been labeled, as shown in Figure 5.

	B	C
1	Bahasa Baku	
2	menggunakan jaket kita ke rumah rumah sakit jiwa	negatif
3	kerusakannya itu loh sangat tidak ada	negatif
4	biarin saja nanti juga besar malu sendiri	negatif
5	bertato saja bangga palingan juga dari pemain karat	negatif
6	kamu kalau mau gaul idung bolang tidak usah di rekam video	negatif
7	sangat tidak nyambung banget	negatif
8	wah sangat keren dan sangat cantik	positif
9	auranya itu loh sangat tidak ada	negatif
10	seperti orang gila bikin kayak seperti itu	negatif
11	sungguh tidak menarik	negatif
12	mengapa dengan dia	negatif
13	untuk apa dia berperilaku seperti itu	negatif
14	saya sangat merinding asli	negatif
15	sungguh menunjukkan membuat seperti itu	negatif
16	suara dan nada terpisah	negatif
17	yang tertawa bukan hanya langit	positif
18	nada sesuai kepercayaan masing-masing	positif
19	beasok akan menjadi mema ini	negatif
20	saya tidak mau kembali kesini lagi	negatif
21	apa malam malam bikin heboh	negatif
22	ini ceritanya membuat apa	negatif
23	aduh sangat talut kepada adik ini	negatif
24	nanti jika dewasa jangan seperti ini ya dek	negatif
25	dik minuman itu membuat menjadi beban lho	negatif
26	sudah minum jamu buyung upi saja de tidak usah minum itu	positif

Figure 5. Labeled data.

d). Tokenizing

Pada proses ini adalah untuk mengubah teks dari kalimat-kalimat yang kompleks menjadi urutan token atau unit-unit yang lebih sederhana. Tujuannya untuk mempersiapkan data teks agar dapat diproses oleh algoritma analisis sentimen dengan lebih efektif dan akurat. Untuk proses tokenizing, Hasil Tokenizing Naïve bayes dapat dilihat sebagaimana yang ada pada gambar 6, gambar 7 dibawah ini.

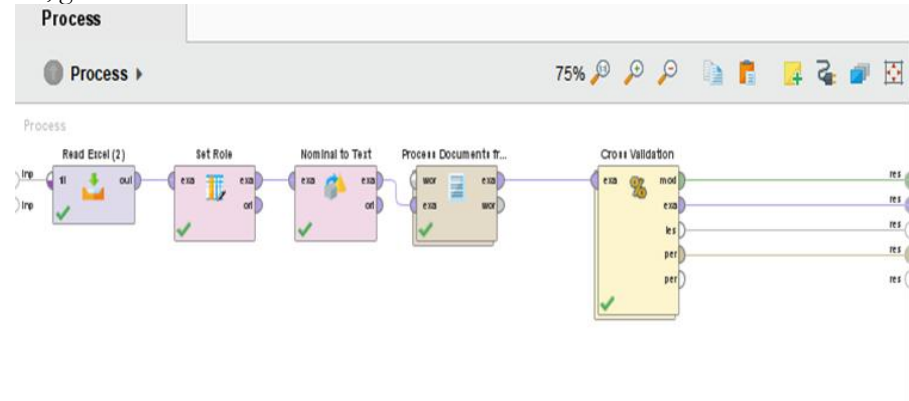


Figure 6. Data Tokenizing Model.

PerformanceVector (Performance (2))

Table View Plot View

accuracy: 47.93% +/- 5.68% (micro average: 47.93%)

	true negatif	true positif	true netral	true negatif	class prec
pred. negatif	52	37	26	2	44.44%
pred. positif	48	97	20	0	58.79%
pred. netral	38	27	43	6	37.72%
pred. negatif	7	4	11	16	42.11%
class recall	35.86%	58.79%	43.00%	66.67%	

Figure 7. Naïve Bayes Tokenizing Results.

Results

The following describes the processing of the rapidminer application as shown in Figure 8.

## PerformanceVector

```

PerformanceVector:
accuracy: 47.93% +/- 5.68% (micro average: 47.93%)
ConfusionMatrix:
True:  negatif      positif netral  negatif
negatif :      52      37      26      2
positif:      48      97      20      0
netral: 38      27      43      6
negatif:      7      4      11      16
    
```

Figure 8. Description of Naïve Bayes Accuracy.

# Simple Distribution

Distribution model for label attribute hasil

Class negatif (0.334)  
1026 distributions

Class positif (0.380)  
1026 distributions

Class netral (0.230)  
1026 distributions

Class negatif (0.055)  
1026 distributions

Figure 9. Description of the Simple Distribution.

The results show a plot view for the Naive Bayes algorithm, as depicted in Figure 10 below:

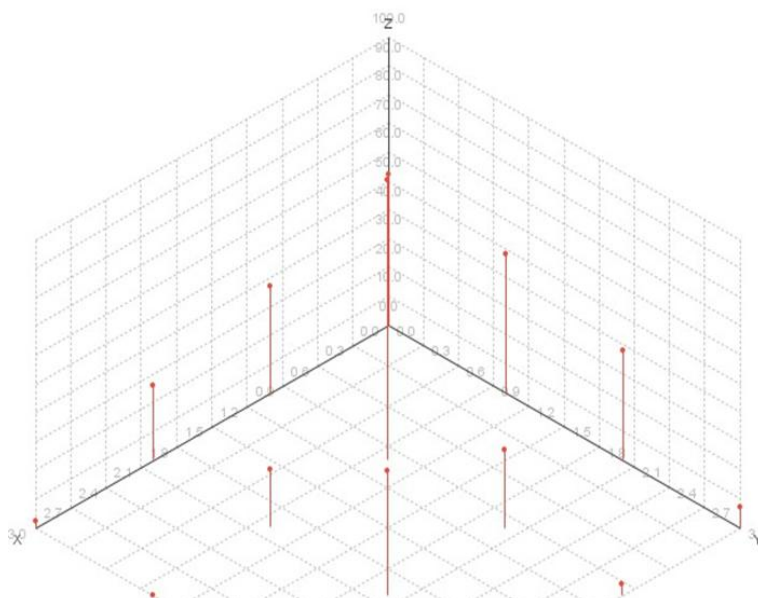


Figure 10. Plot View Naive Bayes.

Table View  Plot View

accuracy: 47.93% +/- 5.68% (micro average: 47.93%)

	true negatif	true positif	true netral	true negatif	class pre
pred. negatif	52	37	26	2	44.44%
pred. positif	48	97	20	0	58.79%
pred. netral	38	27	43	6	37.72%
pred. negatif	7	4	11	16	42.11%
class recall	35.86%	58.79%	43.00%	66.67%	

Figure 11. Table View.

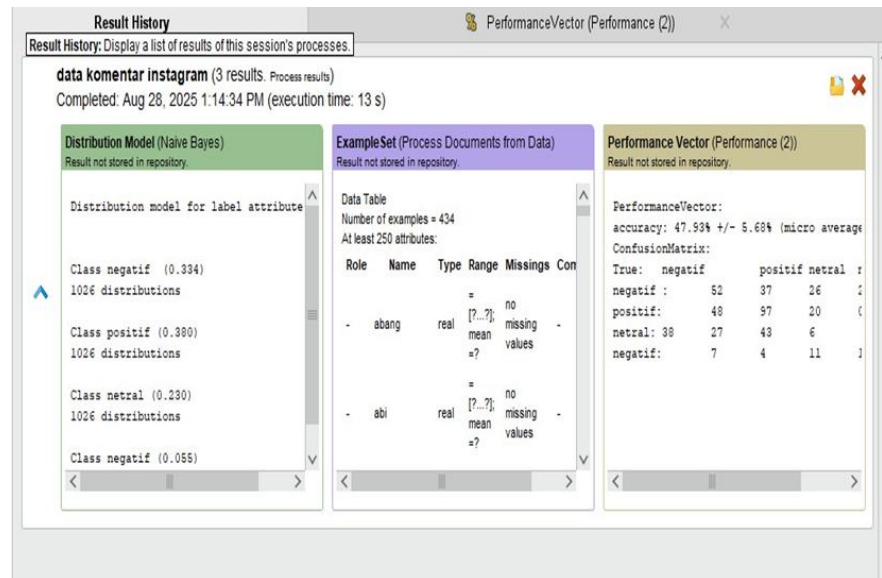


Figure 12. Result History.

Attribute	Parameter	negatif	positif	netral	negatif
abang	mean	0	0.003	0	0
abang	standard deviation	0.001	0.045	0.001	0.001
abi	mean	0	0	0.005	0
abi	standard deviation	0.001	0.001	0.048	0.001
abisan	mean	0	0	0.005	0
abisan	standard deviation	0.001	0.001	0.048	0.001
adakah	mean	0.005	0	0	0
adakah	standard deviation	0.041	0.001	0.001	0.001
adalah	mean	0.005	0.011	0	0
adalah	standard deviation	0.034	0.062	0.001	0.001
adik	mean	0.010	0.007	0.017	0
adik	standard deviation	0.057	0.049	0.083	0.001
adiknva	mean	0.008	0.007	0.007	0

Figure 13. Distribution Table.

Name	Type	Missing	Filter (1,028 / 1,028 attributes):	Search for Attributes
Label <b>hasil</b>	Polynomial	0	Least negatif (24) Most positif (165)	Value
Text <b>text</b>	Text	0	Least yang ter [...] ngit (1) Most bagus sek [...] main (3)	Value
<b>abang</b>	Real	0	Min 0 Max 0.573	Average 0.00
<b>abi</b>	Real	0	Min 0 Max 0.481	Average 0.00
<b>abisan</b>	Real	0	Min 0 Max 0.481	Average 0.00
<b>adakah</b>	Real	0	Min 0 Max 0.410	Average 0.00

Figure 14. Statistic.

## 5. Conclusion and Suggestions

### Conclusion

Based on the sentiment analysis conducted on social media, particularly regarding the use of Instagram by young children, several conclusions can be drawn as follows: a) There are various perspectives and opinions within society regarding the impact of Instagram use by young children. These opinions are divided into three main categories, namely positive, negative, and neutral; b) The use of Instagram by young children is influenced by various factors, including easy access to technology, peer influence, and a lack of supervision and understanding from parents and educators; c) Impact of Instagram Use: The use of Instagram by children can result in both positive and negative effects. Negative impacts include exposure to inappropriate content, the risk of cyberbullying, and disruptions to social and psychological development; d) The Naïve Bayes method has proven effective in sentiment analysis and in classifying public opinions regarding the use of Instagram by young children. This algorithm is capable of providing valuable insights for the development of educational information systems; e) The development of an educational information system based on sentiment analysis can provide comprehensive information and guidance for parents and educators in assisting children to use social media in a healthy and responsible manner.

### Suggestions

Based on the conclusions above, the following suggestions may be considered for future research and practical implementation: a) Prioritize the development of educational information systems that are easily accessible and understandable for parents and educators; b) Include features such as privacy setting guidelines, age-appropriate content education, and strategies to minimize the negative impacts of excessive social media use; c) Parents and educators need to increase supervision and guidance regarding children's use of social media; d) Education concerning the risks and benefits of social media should be provided regularly to children..

## References

- Adelia, S., Milanda, E., Santari, J., Kesuma, D. T., Silvia, E., & Kurniawan, F. (2023). Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi Naive Bayes. *Jurnal Informasi Teknologi Ampere*, 4(3), 254–264. <https://journal-computing.org/index.php/journal-ita/article/view/430>
- Aditya, D. R., Supriyati, E., & Listyorini, T. (2022). Analisis Sentimen Pengguna Twitter Terhadap Rokok Elektrik (Vape) Di Indonesia Menggunakan Metode Naive Bayes. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 7(1), 43–50. <https://doi.org/10.29100/jupi.v7i1.2145>
- Anjani, A., Chamid, A., & Murti, A. (2022). Analisis Sentimen Kaum LGBT pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes. *JTINFO Jurnal Teknik Informatika*, 1(2), 1–8. <https://journal.unisnu.ac.id/JTINFO/article/view/259>
- Devi, A. C., & Fadli, U. M. (2023). Analisis Keputusan Pembelian Produk E-Commerce Shopee Di Kalangan Mahasiswa Universitas Buana Perjuangan Karawang. *Journal of Student Research*, 1(5), 113–123.
- Harahap, S. M., & Kurniawan, R. (2024). Analisis Sentimen Komentar Youtube Terhadap Food Vlogger Dengan Menggunakan Metode Naive Bayes. *MEANS (Media Informasi Analisa Dan Sistem)*, 9(1), 87–96. <https://doi.org/10.54367/means.v9i1.3912>
- Heristian, S., Napiyah, M., & Erawati, W. (2025). Analisis Sentimen Ulasan Pelanggan Menggunakan Algoritma Naive Bayes pada Aplikasi Gojek. *Journal Name*, 5(1), 35–41.
- Laia, E., & Yamin, M. (2023). Penerapan Algoritma Naive Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce. *Media Online*, 4(1), 305–316. <https://doi.org/10.30865/klik.v4i1.1186>

- Marzuki, J., & Sibarani, P. (2024). Implementation of RFM Analysis to Enhance Sales Patterns of Food and Beverages at Bonjour Café and Resto Using the Apriori Algorithm. *Journal Name*, 4(December), 1261–1279.
- Mubarok, D., & Adjani, K. (2025). BIG DATA ANALYTICS DAN MACHINE LEARNING UNTUK MEMPREDIKSI PERILAKU KONSUMEN DI E-COMMERCE. *E-Commerce*, 8(1), 159–167.
- Pratidina, N. D., & Mitha, J. (2023). Dampak Penggunaan Media Sosial terhadap Interaksi Sosial Masyarakat: Studi Literature. *Jurnal Ilmiah Universitas Batanghari Jambi*, 23(1), 810. <https://doi.org/10.33087/jiubj.v23i1.3083>
- Rosidi, R. P. M., & Setiawan, K. (2024). Implementasi Algoritma Naïve Bayes Terhadap Data Penjualan untuk Mengetahui Pola Pembelian Konsumen pada Kantin. *Jurnal Indonesia Manajemen Informatika Dan Komunikasi*, 5(1), 120–126. <https://doi.org/10.35870/jimik.v5i1.407>
- Sanjaya, R., Tohidi, E., & Wahyudi, E. (2024). Media Sosial Twitter Menggunakan Algoritma Naïve Bayes. *Journal Name*, 8(1), 507–514.
- Setiawan, W. D., & Widyaningrum, D. (2024). Komparasi metode algoritma klasifikasi pada analisis sentimen komentar cyberbullying di instagram. *Tekinkom*, 7, 466–472. <https://doi.org/10.37600/tekinkom.v7i1.1392>
- Susanti, N. A., Walid, M., & Hoiriyah, H. (2022). Klasifikasi Data Tweet Ujaran Kebencian Di Media Sosial Menggunakan Naive Bayes Classifier. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 538–543. <https://doi.org/10.36040/jati.v6i2.5174>
- Ula, M., & Fachrurrazi, S. (2023). Analisis Sentimen Cyberbullying pada Media Sosial Twitter menggunakan Metode Support Vector Machine dan Naïve Bayes Classifier. *TECHSI - Jurnal Teknik Informatika*, 14(2), 91. <https://doi.org/10.29103/techsi.v14i2.12103>
- Wahyuni, W. (2022). Analisis Sentimen terhadap Opini Feminisme Menggunakan Metode Naive Bayes. *Jurnal Informatika Ekonomi Bisnis*, 4, 148–153. <https://doi.org/10.37034/infeb.v4i4.162>
- Zamzami, F., Hidayat, R., & Fathonah, R. (2024). Penerapan Algoritma Naive Bayes Classifier Untuk Analisis Sentimen Komentar Twitter Proyek Pembangunan Ikn. *Faktor Exacta*, 17(1), 47–57. <https://doi.org/10.30998/faktorexacta.v17i1.22265>
- Zuraiyah, T. A., Mulyati, M. M., & Harahap, G. H. F. (2023). Perbandingan Metode Naïve Bayes, Support Vector Machine Dan Recurrent Neural Network Pada Analisis Sentimen Ulasan Produk E-Commerce. *Multitek Indonesia*, 17(1), 27–43. <https://doi.org/10.24269/mtkind.v17i1.7092>
- Zusrotun, O. P., Murti, A. C., & Fiati, R. (2022). Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes. *Jurnal Nasional Pendidikan Teknik Informatika*, 11(3), 310–319. <https://doi.org/10.23887/janapati.v11i3.49160>