Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3408

Submitted : Jan 6, 2024
Accepted : Jan 13, 2024
Published : Jan 16, 2024

Sentiment Analysis on Cyanide Case After 'Ice Cold' Aired with NLP Method using Naïve Bayes Algorithm

Rahmatika Hizria[1)*], Sarwadi[2)], Rabiatul Adawiyah Hasibuan [3)], Ramadhani Ritonga[4)], Rika Rosnelly [5)]
1,2,3,4,5) Potensi Utama University, Indonesia
[1)]oshien.zain@gmail.com, [2)]sarwadi69@gmail.com, [3)] rabiatuladawiyahhsb17@gmail.com, [4)] ritongaramadani8@gmail.com, [5)]rika@potensi-utama.ac.id

ABSTRACT

Information technology is developing increasingly rapidly, and the reach of the Internet has expanded even to remote areas. The public increasingly uses social media as a source of information that discusses all aspects of people's lives. Social media has a vital role for most people, one of which is the news of the cyanide coffee case. The Cyanide Coffee case was discussed again by netizens after Netflix raised this case in a documentary film entitled Ice Cold, which made the public even more convinced of the irregularities of the case. Based on this, sentiment analysis is needed to extract comments to obtain public opinion information. The sentiment analysis aims to create a sentiment model to determine public comments on this case. Therefore, this research was conducted to find out and classify public sentiment on the Cyanide Coffee Case using the Natural Language Processing (NLP) method, which is a text preprocessing process followed by the tokenization stage. Data filtering was used using Indonesian Stopwords, and then normalization was continued using Porter Stemmer. In this study, data collection was carried out based on public comments on Ice Cold shows on the TikTok platform using TikTok Comments Scraper. The test results show that the classification using naïve Bayes obtained the results of 22 negative comments, 4052 neutral comments and 34 positive comments. The classification results of this study are 87% accuracy, 97.6% precision, 87% recall, and 91.9% F-Score.

Keywords: Natural Language Processing; Sentiment Analysis; Jessica Wongso; Cyanide Coffee; Ice Cold; Netflix; Naïve Bayes;

INTRODUCTION

Information technology is developing more rapidly, and the reach of the Internet has expanded even to remote areas. Social media is increasingly being used by the public as a source of information in this millennial era, of course, discussing all aspects of life, from social, cultural, economic, and criminal to the community's lifestyle. Social media has a vital role for most people, one of which is the news of the cyanide coffee case.

As we know, the role of social media is vital to raise an issue, make it viral, and get the wider community's attention, with related institutions resolving these issues more quickly.

The Jessica Wongso Cyanide Coffee Case has stolen the spotlight again since it was aired as a documentary on Netflix titled Ice Cold. Since being found guilty of premeditated murder in the death of Wayan Mirna Salihin 7 years ago, Jessica Kumala Wongso has become a hot topic of conversation on social media. Many people's speculations changed when the documentary was aired.

It is known that in January 2016, Indonesian people were shocked by the death of a woman named Wayan Mirna Salihin, who died after drinking Vietnamese Iced Coffee at Olivier Cafe, Grand Indonesia Mall Jakarta, with two friends, Hani and Jessica. Mirna's death was allegedly caused by a corrosive substance found in the coffee she drank.

The news about the case of Mirna's death after drinking coffee containing cyanide went viral and became a trending topic throughout Indonesia; the media paid more attention to the news of the cyanide coffee case and made it a news broadcast continuously every day.

Seven years after the 20-year sentence was handed down to suspect Jessica Kumala Wongso, Netflix turned the story of the cyanide coffee case into a documentary called Ice Cold. After the documentary aired, there were many social stigmas and changes in people's views towards the defendant, Jessica Kumala Wongso. The change in view affects people's belief that Jessica is Mirna's true killer.

Therefore, this research was conducted to find out and classify public sentiment in the Cyanide Coffee case using

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3408

Submitted : Jan 6, 2024
Accepted : Jan 13, 2024
Published : Jan 16, 2024

the Natural Language Processing method. Sentiment analysis is a computational science that studies public opinion on a topic where there is a process of classifying text data containing opinions, whether positive, negative or neutral. The goal is to find out the opinion of a group of people/public on a particular topic, product, service or agency, where the opinion can be positive, negative, or neutral. Information is taken on a text or text mining, such as sentiment analysis, research by (Liu, 2012). Currently, there are many studies related to sentiment analysis. One of them that is currently trending is sentiment analysis on opinions found on social media, research by (Suryani, Linawati, & Saputra, 2019).

The reason why this research uses the Natural Language Processing method is that it is based on a research journal written by Nico Munasatya and Sendi Novianto with the title "Natural Language Processing for Sentiment Analysis of President Jokowi Using Multi-Layer Perceptron", where the research proves that sentiment analysis is commonly used for opinion mining in the sense of giving an identity/label (Positive, Negative, Neutral) to the data/corpus. NLP (Natural Language Processing) is used to process data/corpus so that it can be understood/understood by machines or can be said to be data preprocessing/cleaning text. The classification text used to process the data/corpus is entered into the classification engine model using the multi-layer perceptron model, producing a prediction with a percentage accuracy of > 90% (better), research by (Munasatya & Novianto, 2020). Sentiment analysis using the Naïve Bayes algorithm. The Naïve Bayes algorithm has also been carried out in research on news comments on Twitter; the research was carried out by classifying tweets containing positive and negative comments and producing an accuracy rate of 55.80%, research by (Pandhu & Diki, 2020).

METHOD

This research uses the orange data mining version 3.36 application for opinion mining. According to Turney, Opinion mining or sentiment analysis is the process of understanding, extracting and processing textual data automatically to obtain sentiment information in an opinion sentence. Sentiment analysis is done to see the opinion or tendency of opinion on a problem or object expressed by someone, whether it tends to have a negative or positive view or opinion, research by (Pisceldo, Adriani, & Manurung, 2009). Opinion mining is done to see the opinion or tendency of opinion on a problem or news topic by a person, whether it tends to be negative, positive or neutral, so that it is hoped that the opinions collected can be helpful information. The information contained in online news is unstructured digital text data information, research by (Pang, Lee, & Vaithyanathan, 2002). The sentiment analysis workflow used in this study can be seen in Picture 1. According to (Lisangan, Gormantara, & Carolus, 2022), After the dataset is collected, then through the process of data preprocessing, feature extraction, and classification using Naive Bayes, the algorithm uses a confusion matrix by paying attention to the accuracy, precision, and recall values.

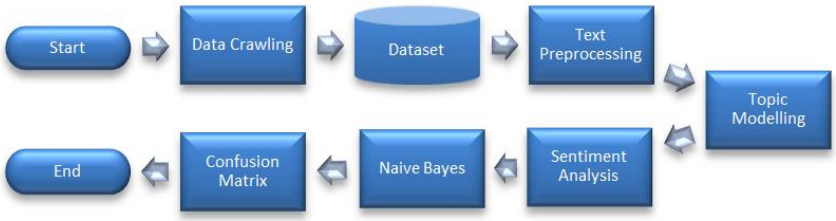

Figure 1. Sentiment Analysis Process

Research Scenario

This research was conducted by utilizing Orange Data Mining Tools. Data from TikTok social media is used as a sample in this study, as shown in Picture 2.

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3408

Submitted : Jan 6, 2024
Accepted : Jan 13, 2024
Published : Jan 16, 2024

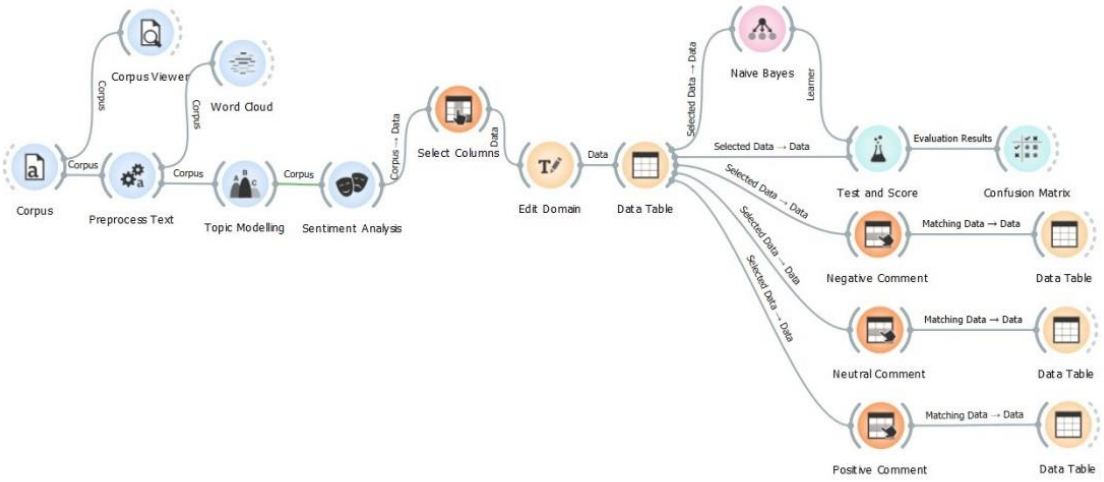

Figure 2. Research Scenario

Data Crawling

In this process, collecting comments and test annotations after Ice Cold is aired, including sarcastic ones, and sentences contained in public comments on the Cyanide Coffee case on the TikTok platform are scraped using the TikTok Comments Scraper tool by entering the keyword searched Jessica Wongso. The dataset collected was 3774 comments in October 2023.

No	createTimeISO	diggCount	replyCommentTotal	text	uniqueId
1	2023-10-07T06:31:02.000Z	4	0	save jessica	rani_2930
2	2023-10-07T06:27:18.000Z	1	0	Film Dokumenter Ice Cold	erjiikz
3	2023-11-14T16:20:43.000Z	0	0	JusticeforJessica	haimin_abadi
4	2023-10-14T08:37:53.000Z	0	0	netflixid	m.galihfitra
5	2023-10-06T15:49:30.000Z	0	0	off Justice jesica	mozza.official2
6	2023-10-14T14:09:12.000Z	0	0	justice jessica	rani_2930
7	2023-10-03T08:58:33.000Z	614	0	emang boleh Jaksa se plot twist itu pak	turuu_____
8	2023-10-03T02:12:20.000Z	5065	24	jaksa Sambo sih bos	purirahayu901
9	2023-10-02T13:55:25.000Z	1964	12	bapak jaksa perlu dipertanyakan yah	elzain101495
10	2023-10-02T13:22:57.000Z	929	2	Bapaknya perlu di pertanyakan.	vvnxp
11	2023-10-05T18:30:24.000Z	618	7	Terima kasih Netflix sudah menayangkan film dokumenter kasus ini, smoga ini jalannya Jesica mendapatkan keadilan lewat Viral lagi. Netizen mari bekerj	xnataj
12	2023-10-02T15:24:51.000Z	4202	45	Katanya ini Jaksa andalan Ferdy sambo	0nepeacexxx
13	2023-10-04T06:02:06.000Z	133	1	di tunggu dokumenter brigadir J	seblakenak84
14	2023-10-03T07:46:13.000Z	167	3	smua pasti akan terbuka siapa yg bermain dan jd sutradaranya	astanto72
15	2023-10-02T14:23:34.000Z	409	9	sayangnya di th itu blm ada ti tok ya	althalia1792
16	2023-10-03T05:43:26.000Z	72	0	bapaknya dan suami Mirna diperiksa kembali untuk mencari kebenaran	nana.shofi
17	2023-10-03T06:08:34.000Z	3000	156	apakah kasus ini yg diketahui oleh Yoshua hingga dia di bunuh?	meeayraa_
18	2023-10-08T04:46:11.000Z	6	0	Jpu shandy berasa bangga amat kasih keterangannya euy, sambil senyum2 😊	staywithme.thrift

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3408

Submitted : Jan 6, 2024
Accepted : Jan 13, 2024
Published : Jan 16, 2024

19	2023-10-07T09:26:23.000Z	5	0	ada kejanggalan	auliyahfarahsum
20	2023-10-06T03:15:34.000Z	5	0	agak lain ini jaksa	brightt_sunnyyy

Table 1. Dataset from TikTok

Text Preprocessing

Preprocessing here is a stage to change the structure of a corpus in the form of a collection of text into tokens or words through the tokenization stage; the tokens are processed again through the cleaning process stage, namely case folding to convert text into lowercase letters (lowercase) and the stopwords removal stage so that the token does not repeat the same word and becomes the base word by removing words that have no value, such as "which", "and", "in", "on". Symbols, emoticons, numbers and punctuation marks are cleaned during the cleaning process. In the last stage, normalization is carried out using a Porter stemmer before the corpus is entered into the classification model to normalize shortened or repeated words.

Topic Modeling

At this stage, sentiment clustering identifies topics into categories or themes to understand sentiment variation and distribution.

Topic modelling is one of the techniques in Natural Language Processing (NLP) to analyze text (Ting, Ip, & Tsang, 2011), an algorithm to identify hidden patterns from a set of words using the technique of distributing words in a set of documents. The output of topic modelling is a set of topics consisting of several clusters of words that appear together in the document based on specific patterns, research by (AGGARWAL & ZHAI, 2013).

Naive Bayes Classifier

Naïve Bayes Classifier is one of the machine learning methods that utilize probability and statistical calculations proposed by British scientist Thomas Bayes, namely predicting future probabilities based on previous experience, research by (Jacobi, Atteveldt, & Welbers, 2015). The basis of Naïve Bayes used in programming is the Bayes formula in equation, research by (Br Ginting & Trinanda, 2013). The probability of event A as B or P(A|B) is determined from the probability of B when A or P(B|A), the probability of A or, and the probability of B or P(A) and the probability of B or P(B).

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$ (Jacobi, Atteveldt, & Welbers, 2015)

Confusion Matrix

The Confusion Matrix is used to measure the classification results of the Naïve Bayes Classifier method. Confusion matrix is a method used to calculate accuracy in data mining concepts. Evaluation by producing accuracy, precision and recall values, research by (Pang, Lee, & Vaithyanathan, 2002).

fij		Prediction Class (j)	
		Class (+)	Class (-)
Actual (i)	Class (+)	TP (true Positif)	FN (Falsa Negative)
	Class (-)	FP False Positive)	TN (true Negative)

Classification accuracy is the percentage of correctly classified data records after testing the classification results.
Precision is the proportion of predicted positive cases and true positives in the actual data.
Recall is the proportion of true positive cases that are correctly predicted positive.
Equation:
Recall = TP / (TP + FN)
Precision = TP / (TP + FP)
Accuracy = (TP + TN) / (TP + FP + FN + TN), research by (Zunic, Corcoran, & Spasic, 2020).

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3408

Submitted : Jan 6, 2024
Accepted : Jan 13, 2024
Published : Jan 16, 2024

RESULT AND DISCUSSION

Based on the dataset, tokens were obtained from the dataset, which are 37,351 tokens with 7,533 types of words. From the transformation process, we obtained 37,300 tokens with 6,560 types of words. After the tokenization process, there were 32,589 tokens with 6,129 types of words. After the filtering process using Indonesian stopwords and Regexp 22,066 tokens with 5,702 types of words. After the normalization process with Porter stemmer, 22,066 tokens with 5,648 types of words were obtained.

The word cloud output shows ten tokens with the highest weight:

Figure 3. Preprocessing results in the word cloud

After obtaining tokens, topic modelling using Latent Semantic Indexing is carried out so that 10 topics are obtained. Using these 10 topics, sentiment analysis uses multi-language Indonesian to get sentiment values (0, <0, >0).

Furthermore, sentiment value normalization is carried out where <0 is a negative category, 0 is a neutral category, and>0 is a positive category.

From the results of this sentiment analysis, there are 10 features and one target (10 topics and sentiment value categories).

The test results on sentiment analysis that have been built obtained that the test results using the naïve Bayes classifier algorithm provide classification test results with an accuracy of 87%, precision of 97.6%, recall of 87% and F-Score of 91.9%.

Model	AUC	CA	F1	Prec	Recall
Naïve Bayes	0.622	0.869	0.917	0.975	0.869

Figure 4. Testing Results with Naïve Bayes Algorithm

The test results of this study show that classification using naïve Bayes obtained the results of 22 negative comments, 4052 neutral comments and 34 positive comments, as shown in Picture 4 below:

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3408

Submitted : Jan 6, 2024
Accepted : Jan 13, 2024
Published : Jan 16, 2024

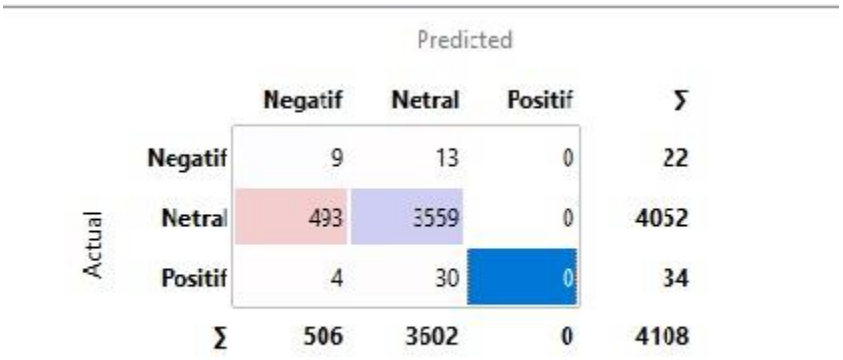

Figure 4. Results with Naïve Bayes Algorithm

CONCLUSION

Based on the research that has been done, it is concluded that AUC (Area Under ROC Curve) has the lowest value due to the imbalance in the classification results using the naïve Bayes classifier algorithm because it produces better performance accuracy in the application of the classification process

REFERENCES

Aggarwal, C. C., & Zhai, C. (2013). Mining text data. *Mining Text Data 9781461432*, 1–522. doi: 10.1007/978-1-4614-3223-4.

Br Ginting, S. L., & Trinanda, R. P. (2013). Teknik Data Mining Menggunakan Metode Bayes Classifier untuk Optimalisasi Pencarian padAplikasi Perpustakaan (Studi Kasus: Perpustakaan Universitas Pasundan – Bandung). *Jurnal Teknologi dan Informasi*, DOI: 10.34010/jati.v3i2.794.

Jacobi, C., Atteveldt, v. W., & Welbers, K. (2015). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 89-106. https://doi.org/10.1080/21670811.2015.1093271.

Lisangan, E. A., Gormantara, A., & Carolus, R. Y. (2022). Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia. *KONSTELASI Konvergensi Teknologi dan Sistem Informasi*, 2(1).

Liu, B. (2012). Morgan & Claypool Publishers.

Munasatya, N., & Novianto, S. (2020). Natural Language Processing untuk Sentimen Analisis Presiden Jokowi Menggunakan Multi Layer Perceptron. *Techno Com*, 19(3):237-244.

Pandhu, A., & Diki, W. (2020). Analisa sentimen dan Klasifikasi Komentar Positif Pada Twitter dengan Naïve Bayes Classification. *BRITech (Jurnal Ilmiah Ilmu Komputer, Sains dan Teknologi Terapan)*, 1(2). 32–40.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Association for Computational Linguistics*, 79–86.

Pisceldo, F., Adriani, M., & Manurung, R. (2009). Probabilistic Part Of Speech Tagging for Bahasa Indonesia. *Third International MALINDO Workshop, colocated Event ACLIJCNLP*.

Suryani, N. S., Linawati, & Saputra, K. O. (2019). Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia. *Majalah Ilmiah Teknologi Elektro*, 22.

Ting, J. S., Ip, W. H., & Tsang, A. H. (2011). Is Naïve Bayes a Good Classifier for Document Classification? *International Journal of Software Engineering and Its Applications*, 5(3).

Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment Analysis in Health and Well-Being: Systematic Review. *JMIR Medical Informatics*, 8(1) 1-22. doi : 10.2196/16023.

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).