Journal of Student Research Exploration

https://shmpublisher.com/index.php/josre
p-ISSN 2746-7686 | e-ISSN 2746-0991

Comparative sentiment analysis on starbucks boycott related to Israel's genocide using svm, naïve bayes, and knn methods

Luthfiyyah Wahyu Nurfarida[1], Budi Prasetyo
[1,2]Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received July 12, 2024
Revised October 3, 2025
Accepted October 3, 2025

Keywords:

Sentiment analysis
Starbucks
Boycott
Israel genocide

ABSTRACT

The #BoycottStarbucks movement emerged in response to allegations of Starbucks' involvement in supporting the genocide carried out by Israel in Palestine. This research aims to analyze public sentiment towards Starbucks regarding its alleged involvement in the Israeli genocide. Data was collected through crawling English-language comments on a YouTube channel named Firstpost, which discussed the boycott of Starbucks. Three classification methods used were Naive Bayes, KNN, and SVM. The research findings indicate that the SVM model had the highest accuracy (82%) compared to Naive Bayes (75%) and KNN (56%). The distribution of sentiment in the comment data was dominated by neutral sentiment (44.5%), followed by positive sentiment (37.3%) and negative sentiment (18.2%). However, positive comments were also significant, indicating support or interest in the boycott movement. The hypotheses proposed supported the finding that public sentiment towards Starbucks tends to be negative and supportive of the boycott movement regarding its alleged involvement in the Israeli genocide. The research also concludes that various public opinions are reflected in neutral, positive, and negative comments, indicating that the boycott of Starbucks still attracts significant attention and discussion.

1. Introduction

[1] ***Corresponding Author:***
Luthfiyyah Wahyu Nurfarida,
Department of Computer Science,
Universitas Negeri Semarang, Indonesia
Email: workfiyy@gmail.com
DOI: https://doi.org/10.52465/josre.v3i2.421

In recent years, the #BoycottStarbucks movement has acquired global traction, advocating for a boycott of the well-known fast food giant Starbucks. The move was in response to allegations that Starbucks was complicit in the Israeli genocide in Palestine [1]. Within months, the boycott effort against Israel had moved from liberal campuses and protest marches to the center of attention in Congress [2]. This movement demonstrates how quickly this issue gathers traction and becomes a major legislative priority. The claims stem from Starbucks' involvement in Israeli enterprises involved in the establishment of illegal settlements in Palestinian land, as well as its backing for Israel's military. Western support for Israel stems from common values, national interests, and history, particularly in colonial-colonial countries such as the United States, Canada, and Australia. In the United States, Israel and the United States are viewed as fighting for democracy in challenging geopolitical situations. Cultural and religious links reinforce this closeness. However, as human rights awareness grew, public opinions began to shift. The younger generation is increasingly doubting its unconditional support for Israel, taking into account the Palestinian plight and denouncing Israel's human rights violations. This perspective is crucial in assessing the evolution of political and cultural support across time in a global environment [3]. Previous study on the Starbucks boycott centered on the movement's economic and political dimensions. This sentimental analysis is critical for understanding public perception of this issue and how it influences consumer behavior. The study's goal is to examine public perceptions of Starbucks in light of its suspected complicity in the Israeli genocide.

Western support for Israel stems from common values, national interests, and history, particularly in colonial-colonial countries such as the United States, Canada, and Australia. In the United States, Israel and the United States are viewed as fighting for democracy in challenging geopolitical situations. Cultural and religious links reinforce this closeness. However, as human rights awareness grew, public opinions began to shift. The younger generation is increasingly doubting its unconditional support for Israel, taking into account the Palestinian plight and denouncing Israel's human rights violations. This perspective is crucial in assessing the evolution of political and cultural support across time in a global environment [3].

Previous study on the Starbucks boycott centered on the movement's economic and political dimensions. This sentimental analysis is critical for understanding public perception of this issue and how it influences consumer behavior [4].

The study's goal is to examine public perceptions of Starbucks in light of its suspected complicity in the Israeli genocide and to know what is the most accurate way to quantify the emotive impact of the Starbucks boycott. The goal of this study is to better comprehend public opinion toward Starbucks in light of its claimed role in the Israeli genocide, as well as to determine which approaches are most accurate for analyzing public sentiment toward the topic.

A hypothesis is a statement formed based on initial observations and can be tested through research or experiments [5]. The hypothesis proposed by the author

includes (1) Public sentiment towards Starbucks would be negative in connection with his alleged involvement in the Israeli genocide. (2) The public will support the Starbucks Boycott Movement in connection with its involvement in the Israeli genocide. (3) The support vector machine (SVM) method will have the highest degree of accuracy compared to other methods.

2. Method

In this research, it uses qualitative methods that organize explorative approaches. The data was collected through crawling English-language Youtube comments on a channel called Firstpost that discussed the Starbucks boycott. Here's the stages of data processing that I'm using in this sentimental analysis.

Crawling Data

YouTube comment data is collected using crawling techniques with the help of the library googleapiclient.discovery, pandas, and textblob in Python. This crawl process allows researchers to access and extract comment data automatically from the YouTube videos that are the subject of the research.

Text Mining

Text mining is part of language processing technology. In research, this method is used to understand materials such as case studies, personal stories, interviews, and visual texts. This material can describe everyday moments and problems, as well as meaning in a person's life [6]. Here's a table of data crawling results.

Table 1. Results of crawling youtube comments

	Text
0	I support israel. But I stopped supporting sta...
1	Msia economy down is main factor .....
2	Coffee jobs that supposed to be STARTER jobs, ...
3	I just bought a green tea cream, it tastes del...
4	In America Starbucks is anti-jewish. Starbucks...
...	...
3575	Second comment
3576	A granda problem lol

Data Processing

Data that has been successfully crawled or collected is then processed through several stages.

Cleaning Data

Data that has been successfully crawled will be pre-cleaned at this stage to clear unnecessary characters, remove empty text, as well as include changing characters to small letters, filtering, and replacing irrelevant or unneeded elements. In this process, I used the Rapid Miner tools to clean up the data. The Rapid Miner application is used for training, research, as well as role in rapid prototyping, and application development, and supports all steps of the machine learning process including data preparation, visualization, validation and optimization processes [7], [8]. Here's the pre-processing phase I'm using.

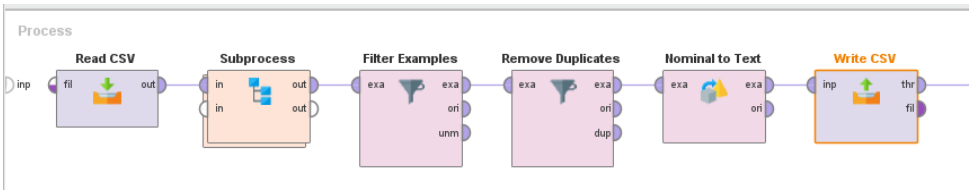

Figure 1. Data processing

The Rapid Miner application will read the CSV file from the crawling data that has been done. After the document is read, the data will be processed by removing the unnecessary URLs and characters, and will be trimmed, this operator will remove spaces in the front and back of the selected nominal attribute values. After that, make sure the document does not have empty data with the example filter operator to filter and then remove the data considered as missing value. After that, the data is cleared from redundancy to eliminate duplication of data. To ensure that the data can be processed, the nominal type data is converted to text form. The purpose of this data cleaning phase is to ensure that the data to be analyzed is in clean condition and ready for use.

Table 2. Clean text data result from rapid miner

Text
I support israel But I stopped supporting starbucks years ago when they decided to creat jobs for muslim quot refugees quot over americans
Msia economy down is main factor
Coffee jobs that supposed to be STARTER jobs not life supporting jobs You want better money Get a freaking skill that exceeds pouring coffee and and mixing fancy foam and or get another job
I just bought a green tea cream it tastes delicious 100 worth the money
In America Starbucks is anti jewish Starbucks

Preprocessing

After data cleaning through Rapid Miner, the data is processed again with Google Colab. This processing aims to ensure that the analysis process is ready to be performed against documents and data. The first stage, Transform case or change the whole word into a small letter to match the content of the data. After the whole data uses small letters, the removal of unnecessary characters such as symbols,

reading marks, and characters other than letters and numbers. The following table lists the output of the data after processing.

Table 3. Cleaned text results

	text	clean_text
0	i support israel but i stopped supporting sta...	i support israel but i stopped supporting sta...
1	msia economy down is main factor	msia economy down is main factor
2	coffee jobs that supposed to be starter jobs ...	coffee jobs that supposed to be starter jobs ...
3	i just bought a green tea cream it tastes del...	i just bought a green tea cream it tastes del...
4	in america starbucks is anti jewish starbucks	in america starbucks is anti jewish starbucks

The comment data is then analyzed using the TextBlob library. Here is a table of the results of the successful sentiment analysis.

Table 4. Sentiment analysis results on text

	Text	Clean_text	Sentiment	Sentiment_class
0	i support israel but i stopped supporting sta...	i support israel but i stopped supporting sta...	0.250000	Positive
1	msia economy down is main factor	msia economy down is main factor	0.005556	Positive
2	coffee jobs that supposed to be starter jobs ...	coffee jobs that supposed to be starter jobs ...	0.375000	Positive
3	i just bought a green tea cream it tastes del...	i just bought a green tea cream it tastes del...	0.366667	Positive
4	in america starbucks is anti jewish starbucks	in america starbucks is anti jewish starbucks	0.000000	Neutral
...	...	...	...	...
3003	as from now i will ask anyone i know to have a...	as from now i will ask anyone i know to have a...	0.000000	Neutral
3004	starbucks coffee boycott here philippines	starbucks coffee boycott here philippines	0.000000	Neutral
3005	second comment	second comment	0.000000	Neutral
3006	a granda problem lol	a granda problem lol	0.800000	Positive

After giving sentiment, the comment data goes through a tokenization process to break down the words in the comment into individual tokens. In addition, a stopwording process is carried out to remove words and predicates that have no meaning in this research, as well as stemming to change the words into their basic form.

Table 5. Stemmed data results

	Clean_Text	Sentiment	Sentiment_Class	Stemmed_Text
0	i support israel but i stopped supporting star...	0.250000	Positive	i support israel but i stopped supporting star...
1	msia economy down is main factor	0.005556	Positive	msia economy down is main factor
2	coffee jobs that supposed to be starter jobs n...	0.375000	Positive	coffee jobs that supposed to be starter jobs n...
3	i just bought a green tea cream it tastes deli...	0.366667	Positive	i just bought a green tea cream it tastes deli...
4	in america starbucks is anti jewish starbucks	0.000000	Neutral	in america starbucks is anti jewish starbucks

Data and Samples

The study used 3577 YouTube comment data. After a process of data cleaning, a net data of 3007 data was obtained, which was then divided into two sets, including Test data of 20% of the total data (601 comment data) is used as training data to build a model, and Training data of 80% of total data (2406 comment data), used as test data to test the performance of a model that has been built.

Data Attributes

The collected data has two main attributes, namely the full_text column that stores the text of the collected YouTube comments and the sentiment column to store the sentiment analysis results of each comment, which is classified as negative or positive.

Classification Methods

The research used three classification methods to analyze YouTube's comments sentiment, including:

Naive Bayes

This method assumes that all features used for classification are independent of each other. The advantage of this method is its simplicity and ease of implementation. However, the performance of this method may decrease if there is a high correlation feature [9], [10].

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \quad (1)$$

Where:

- $P(C|X)$ is a class C probability of data X
- $P(X|C)$ is the probability of data X against class C.
- $P\,(C)$ is a priori probability of class C.
- $P\,(X)$ is the probability of the X data.

In the context of text classification, the Naive Bayes multinomial assumption is usually used [11]. Where xi is a feature (word) in the X data.

$$P(C|X) \propto P(C) \cdot \prod_{i=1}^{n} P(xi \mid C) \quad (2)$$

K-Nearest Neighbors (KNN)

This method classifies new data based on the nearest data in the feature space using the value K [12]. The KNN method is easy to interpret and does not require any assumptions about the data distribution. However, performance depends heavily on the selection of k values and the number of features used [13].

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (3)$$

Where:

- x and y is two data points in a feature space.
- x_i and y_i is a component of the data point x and y in i dimention

Support Vector Machine (SVM)

SVM is a machine learning method that operates on the principle of Structural Risk Minimization (SRM) with the aim of finding the best hyperplane that can separate two classes in an input space [14]. SVM is known to have strong and reliable performance in dealing with complex data, although it is more complex to implement and interpret.

$$K(x_i x_j) = \exp\left(-\gamma \left|\left|x_{i-} x_j\right|\right|^2\right) \quad (4)$$

Where γ is the kernel parameter that controls the influence distance from one data sample to another.

Model Evaluations

To measure the performance of the model, some commonly used evaluation metrics can be used including accuracy, precision, recall, F-1 Score.

Accuracy

Accuracy is the parameter used to evaluate the systems built in this study. Accuracy is often referred to as the level of correspondence between the predicted value and the actual value. The accuracy formula can be seen below [15]:

$$\frac{The\ number\ of\ predictions\ is\ correct}{Total\ predictions}$$

Precision

Precision is the number of users classified in a class against the total of users in that class [16]. Precision often is meant as the degree of accuracy between an information a user requests and the answers the system gives. Here's the formula of precision:

$$\frac{The\ number\ of\ predictions\ is\ correct}{Total\ predictions}$$

Recall

Recall is the ratio of users who are well-classified in a class to the number of users in that class [15]. Recall also is an index of the success of the system in retrieving the information sought [17]. The recall formula can be expressed as follows:

$$\frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negative\ (FN)}$$

F1 Score

$$\frac{Presisi\ .\ Recall}{Presisi\ +\ Recall}$$

The use of these metrics helps in understanding the performance of each method of classification and choosing the best method based on the results of the evaluation.

3. Results and Discussion

Public Sentiment Towards the Boycott Starbucks Issue

Based on the data processing that the researchers have done, here is a table of sentiment distribution results in the YouTube comment data.

Table 6. Sentiment analysis of comments results

Sentimen	Jumlah	Persentase
Negative	555	18.46%
Neutral	1403	46.66%
Positive	1049	34.89%
Total	3577	100.00%

According to the table, comments with neutral sentiment dominated (46.66%), followed by positive comments (34.89%) and negative comments (18.46%). It shows that the majority of YouTube viewers are still explicitly expressing support for the Starbucks boycott movement. However, there may be some discussion or debate going on in the comment column.

Public Sentiment Towards the Boycott Starbucks Issue

Word cloud is a visual representation of a text that indicates the frequency or importance of certain words in the text. Based on the processing and analysis of

commentary data that the author has made on the issue of boycott by the world community against Starbucks as a form of protest against the company's involvement in the Israeli-Palestinian genocide issue, the following are the words that appear in this study.

Figure 2. Wordcloud based on net data

Evaluation of Classification Model

Support Vector Machine

Table 7. Analysis results using the svm method

	Precision	Recall	F1-Score	Support
Negative	0.63	0.57	0.6	110
Neutral	0.88	0.89	0.89	288
Positive	0.81	0.84	0.82	204
Accuracy			0.82	602
Macro avg	0.77	0.77	0.77	602
Weighted avg	0.81	0.82	0.81	602

Based on the results of the SVM analysis shown in the picture, it can be concluded that the analysis using SVM has performed well in classifying comment sentiment into negative, neutral, and positive. This is shown with an 82% accuracy value that belongs to a fairly high level. However, the performance of SVM in the negative class (negative sentiment) needs to be improved due to low precision and recall values. (0.63 dan 0.14).

K-Nearest Neighbor

Table 8. Analysis results using the kkn method

	Precision	Recall	F1-Score	Support
Negative	1	0.07	0.14	110
Neutral	0.52	0.99	0.68	288
Positive	0.9	0.22	0.35	204

Accuracy			0.56	602
Macro avg	0.81	0.43	0.39	602
Weighted avg	0.74	0.56	0.47	602

The KNN model has the lowest accuracy (56%) among the three models. Besides, the KNN performance in the negative class is also low (precision 0.24 dan recall 0.05).

Naïve Bayes

Table 9. Analysis results using the naïve bayes method

	Precision	Recall	F1-Score	Support
Negative	0.74	0.45	0.56	110
Neutral	0.84	0.79	0.81	288
Positive	0.67	0.86	0.76	204
Accuracy			0.75	602
Macro avg	0.75	0.70	0.71	602
Weighted avg	0.76	0.75	0.75	602

Based on Naive Bayes' analysis, it can be concluded that this model is quite effective in classifying commentary sentiment into negative, neutral, and positive. It's shown with an accuracy of 75% which belongs to quite high. The sentimental diversity of the comment data affects the performance of this method model to identify accurate sentimental patterns.

SVM is capable of handling non-linear data, which means that SVM can be used to classify data that has nonlinear patterns. This is an advantage to SVM over other classification algorithms such as Naive Bayes, which can only handle linear data. Based on the evaluation, the SVM model showed the best performance in classifying YouTube's overall sentimental comments. However, SVM's ability to identify negative sentiments still needs to be improved. The SVM model is one of the most effective classification algorithms for comment sentiment analysis. By choosing the right kernel, improving the data quality, and trying other algorithms, the performance of the SVM model in comment sentiment classification can be improved.

Although the Naive Bayes model has a lower overall accuracy than the SVM, it has a good performance in identifying positive sentiment. This method will be useful for further research that focuses more on identifying supporters of the Starbucks boycott movement. On the contrary, the KNN model has the lowest performance

among the three models. This suggests that the KNN technique is less effective in dealing with the Youtube comment data in this study.

4. Conclusion

The research managed to conduct a sentimental analysis of English-language YouTube comments related to the Starbucks boycott. The distribution of sentiment in comment data is dominated by neutral sentiment, followed by positive and negative sentiment. The SVM model shows best performance in classifying sentiment, while the Naive Bayes model has good performance in identifying positive sentiment.

Based on the data presented, it can be concluded that the majority of comments related to the Starbucks boycott movement tended to be neutral (46.66%), followed by positive comments (34.89%) and negative comments (18.46%).

The public movement against this Starbucks boycott movement still seems to be attracting attention and sparking debate, as on YouTube. Overall, sentimental analysis suggests that Starbucks boycott moves are still relevant and trigger a lot of talk.

REFERENCES

[1] M. Khoiruman dan A. Wariat, "Analisa Motivasi Boikot (Boycott Motivation) Terhadap Produk Mc Donald Di Surakarta Pasca Serangan Israel Ke Palestina," *J. Manajemen, Bisnis dan Pendidik.*, vol. 10, no. 2, hal. 247–257, 2023, [Daring]. Tersedia pada: https://e-journal.stie-aub.ac.id/index.php/excellent.

[2] M. W. David M. Halbfinger dan S. Erlanger, "Is B.D.S. Anti-Semitic? A Closer Look at the Boycott Israel Campaign," *The New York Times*, 2019. https://www.nytimes.com/2019/07/27/world/middleeast/bds-israel-boycott-antisemitic.html.

[3] J. M. Hitchcock, "A Rhetorical Frame Analysis of Palestinian-Led Boycott, Divestment, Sanctions (BDS) Movement Discourse," 2020, doi: 10.25777/gq1b-4m33.

[4] S. P. Pandey dan M. Y. Ahmad, "An Analysis of Consumer Motivations for Boycott," *J. Univ. Shanghai Sci. Technol.*, vol. 22, no. 10, hal. 2016–2038, 2020.

[5] A. Kurniawan, A. Hidayat, dan B. T. Andika, "Pengaruh Persepsi Kegunaan dan Persepsi Kemudahan terhadap Kepercayaan dan Kepuasan Konsumen pada Aplikasi Dompet Digital," *Indones. J. Econ. Business, Accounting, Manag.*, vol. 01, no. 03, hal. 71–84, 2023.

[6] L. Wang, Y. Huang, dan M. K. Omar, "Analysis of Blended Learning Model Application Using Text Mining Method State of Art," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 01, hal. 172–187, 2021, doi: https://doi.org/10.3991/ijet.v16i01.19823 Lin.

[7] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, dan N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. April, hal. 47–58, 2021, doi: http://dx.doi.org/10.28932/jutisi.v7i1.3216.

[8] M. R. Ningsih dan Jumanto, "Sentiment Analysis on SocialMedia Using TF-IDF Vectorization and H2O Gradient Boosting for Student Anxiety Detection," *Sci. J. Informatics*, vol. 11, no. 4, hal. 1137–1144, 2024, doi: 10.15294/sji.v11i4.20582.

[9] M. R. Ningsih, "Classification Email Spam using Naive Bayes Algorithm and Chi-Squared Feature Selection," vol. 9, no. 1, hal. 74–87, 2024.

[10] E. W. Sholeha, S. Yunita, R. Hammad, V. C. Hardita, dan Kaharuddin, "Analisis Sentimen Pada

Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor (Sentiment Analysis of Online Travel Agent Using Naïve Bayes," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, hal. 203–208, 2022.

[11] M. R. Maʼarif, "Perbandingan Naïve Bayes Classifier Dan Support Vector Machine Untuk Klasifikasi Judul Artikel," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 2, hal. 90–93, 2016, doi: https://doi.org/10.14421/jiska.2016.12-05.

[12] A. H. Siregar, A. A. Tumanggor, dan A. Rahmadani, "Penerapan K-Nearest Neighbors (KNN) dalam Memprediksi dan Menghitung Akurasi Data Penyakit Stroke," *J. Penelit. Rumpun Ilmu Tek.*, vol. 2, no. 4, hal. 146–154, 2023, doi: https://doi.org/10.55606/juprit.v2i4.3040.

[13] Legito *et al.*, "Implementation K-Nearest Neighbor Algorithm for Sentiment Analysis on Khilafah and Radicalism Issues in Indonesia," *Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, hal. 324–330, 2023, doi: https://doi.org/10.57152/malcom.v3i2.893.

[14] A. S. Nugroho, A. B. Witarto, dan D. Handoko, "Support Vector Machine," *Proceeding Indones. Sci. Meet. Cent. Japan*, 2003, [Daring]. Tersedia pada: https://www.academia.edu/24381027/Support_Vector_Machine_Teori_dan_Aplikasinya_dalam_Bioinformatika_1.

[15] P. Tan dan M. Steinbach, "Introduction to Data Mining Instructor ' s Solution Manual," 2006.

[16] H. S. Ginting, K. M. Lhaksmana, dan D. T. Murdiansyah, "Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes," *e-Proceeding Eng.*, vol. 5, no. 1, hal. 1793–1802, 2018.

[17] S. Nurul, J. Fitriyyah, N. Safriadi, dan E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, hal. 279–285, 2019.