JESICA: Journal of Enhanced Studies in Informatics and Computer Applications
Journal homepage: jesica.itsk-soepraoen.ac.id
e-ISSN: 3046-6938 p-ISSN: 3046-6997
Vol. 3, No. 2, July 2026, Pp. 56-71

A Systematic Review of Machine Learning Approaches for K-Pop Concert Sentiment Analysis on X

Yunia Putri Adifa*[1], Carli Apriansyah Hutagalung[2]
1,2 Department of Computer Science, Faculty of Mathematics and Natural Sciences, State University of Jakarta, Indonesia

***Corresponding author**
E-mail address:
adifayuniaputri@gmail.com

Keywords:
K-pop Concerts, Machine Learning, Sentiment Analysis, Systematic Review, X

Abstract

K-pop concerts in Indonesia generate intensive digital discussion on X/Twitter, yet studies that directly combine concert-related objects, X data, and machine-learning-based sentiment classification remain limited. This study conducts a systematic literature review to map research objects, methods, preprocessing techniques, evaluation results, and research gaps in K-pop sentiment analysis. The selection process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses framework. A total of 5,674 records were collected, 4,618 duplicates were removed, 1,056 records were screened, and 20 articles were included in the final synthesis. The findings show that X is the dominant platform. Previous studies more frequently examine K-pop groups, fandom, the Korean Wave, and cyberbullying rather than direct concert experiences. Naive Bayes remains widely used because it is simple, efficient, and suitable for high-dimensional text data, although Support Vector Machine and transformer-based models often provide stronger performance in specific settings. Classification quality is strongly affected by non-standard language normalization, multilingual content, class balance, feature weighting, and labeling consistency. The main gap is the absence of an Indonesian K-pop concert sentiment-analysis design that combines domain-aware preprocessing, per-class evaluation, and aspect-level interpretation of ticketing, promoters, venues, safety, and audience experience.

1. Introduction

The development of K-pop can no longer be understood only as a musical trend. It operates as a transnational popular-culture ecosystem involving content production, digital distribution, fandom interaction, merchandise consumption, live events, and continuous cross-border communication. Within this ecosystem, concerts are important because they connect an offline entertainment experience with an intensive digital conversation that begins before ticket sales, grows during the event, and continues through post-concert evaluation.

Social media provides a primary space in which fans express support, criticism, disappointment, and personal experience. Research on the success of K-pop shows that digital participation is connected with national image, social-network citizenship behavior, and behavioral intention [1]. Fan discourse is therefore relevant not merely as entertainment commentary but as evidence of how popular culture, identity, community, and communication behavior interact in networked environments.

X/Twitter has characteristics that distinguish it from other platforms. Public text can circulate rapidly, hashtags and mentions organize discussion, and users can respond to an event in real time. K-pop stan communities on Twitter have been described as communities of practice because its members build relationships, exchange knowledges, and coordinate collective activities [2]. These properties make X suitable for studying about ticketing, promoters, fan projects, queues, venue access, safety, artist performance, and audience satisfaction.

In Indonesia, K-pop concerts regularly generate large volumes of conversation because the country has active fan communities and a growing live-entertainment market. Posts do not contain enthusiasm alone. They also include complaints about ticketing systems, prices, crowd control, venue facilities, cancellations, technical problems, and post-event services. For promoters, researchers, and fan communities, these posts can provide systematic evidence of public perception, although their volume makes manual analysis inefficient.

Sentiment analysis offers a computational approach for assigning opinions to positive, negative, or neutral classes. Sentiment and opinion mining examine evaluations, attitudes, and emotions expressed in text [3]. In a concert context, the method can identify whether discussion supports an event, criticizes its organization, or mainly distributes

information. A more detailed analysis can relate polarity to specific aspects such as ticket availability, promoter communication, venue comfort, security, and the quality of the performance.

X data also introduces substantial linguistic challenges. Indonesian, English, and Korean expressions may appear in a single post together with fandom slang, abbreviations, emoji, irony, sarcasm, elongated words, and hashtags. Consequently, cleaning, case folding, tokenization, normalization, stop-word removal, stemming, and term weighting directly affect the representation received by a classifier. An inappropriate preprocessing decision may remove a useful concert aspect or reverse the meaning of a negated expression.

Naive Bayes is frequently used for text classification because it is probabilistic, computationally efficient, and effective with high-dimensional sparse features [4]. It estimates the probability that terms occur in each class and uses those probabilities to predict the label of a new document. Its simplicity makes it a practical baseline for student research, moderate datasets, and studies that require transparent procedures rather than resource-intensive training.

Nevertheless, Naive Bayes assumes conditional independence among features, whereas natural language is highly contextual. Its performance depends on annotation quality, class balance, feature representation, and domain vocabulary. K-pop posts include member names, venue names, ticket categories, promoter accounts, fandom expressions, and multilingual abbreviations that are absent from general dictionaries. A model may therefore misinterpret a supportive fandom phrase as neutral or treat criticism of a promoter as criticism of the artist.

Previous studies have applied Naive Bayes to K-pop, the Korean Wave, cyberbullying, and public opinion on X, while other studies compared Naive Bayes with SVM or adopted CNN, IndoBERT, and XLM-RoBERTa [5], [8] - [14]. A study of Treasure fans at Karnaval Mandiri reported high performance [12], while research on NCT concerts further indicates that concert discourse contains distinctive issues such as information quality, ticket fraud, and technical constraints [5].

A systematic literature review is required to organize these fragmented findings. PRISMA 2020 provides a transparent framework for identification, screening, eligibility assessment, and inclusion [6]. This study therefore maps the development of K-pop sentiment research, compares objects and algorithms, evaluates preprocessing and reporting practices, and identifies the gap surrounding Indonesian K-pop concerts on X. The review is designed to provide both a methodological baseline and a practical direction for future concert-focused analysis.

The purpose of this systematic review is to consolidate fragmented evidence on the research objects, platforms, algorithms, preprocessing choices, and evaluation practices used in K-pop concert sentiment analysis. By distinguishing directly concert-focused studies from methodologically or contextually supporting work, the review provides a reproducible evidence map for researchers, identifies reporting standards for comparable experiments, and offers practitioners a basis for interpreting public responses to ticketing, promoters, venues, safety, performance, and audience experience. These contributions support methodological development in multilingual social-media analytics and evidence-based improvement of concert services.

2. Research Method

This section describes the methodological procedures employed to conduct the systematic literature review. It explains the review design and research questions, the literature search strategy, the inclusion and exclusion criteria, the PRISMA-based study selection process, and the data extraction and analysis procedures. These stages were designed to ensure that the identification, selection, and synthesis of relevant studies were conducted systematically, transparently, and consistently with the objectives of the research.

2.1. Review Design and Research Questions

This study applies a systematic literature review to examine, compare, and synthesize published research rather than collect primary posts from social-media users. The review design makes it possible to identify recurring patterns in objects, platforms, methods, preprocessing, evaluation, contributions, and limitations. The procedure follows a sequence of identification, deduplication, title-and-abstract screening, eligibility assessment, data extraction, and narrative-thematic synthesis.

PRISMA was selected as the reporting framework because it requires the number of records and the reasons for exclusion to be stated at each stage [6]. This structure improves transparency and enables readers to reconstruct how a broad search was reduced to the final group of studies. The approach also reduces the risk that the review becomes an unsystematic collection of summaries.

The analytical scope combines computational and social perspectives. Articles that directly analyze K-pop sentiment on X are treated as the most relevant evidence. Studies of fandom, Korean popular culture, cyberbullying, or

general X sentiment are retained when they contribute to understanding the platform, the language domain, a classification method, or a research limitation.

- (RQ1): How have the research objects, platforms, and methods in K-pop sentiment and social-media studies developed?
- (RQ2): Which preprocessing procedures, classification models, and evaluation strategies have been used, and how do their reported results differ?
- (RQ3): What are the strengths, limitations, and research gaps that support a specific study on sentiment analysis of K-Pop concerts in Indonesia on X?

2.2. Literature Search Strategy

The literature search combined four concept groups: sentiment analysis, K-pop or concerts, X/Twitter or social media, and text-classification methods. The principal Boolean string was ("sentiment analysis" OR "opinion mining") AND ("K-pop" OR "BTS" OR "NCT" OR "Blackpink" OR "Korean Wave" OR "K-pop concert" OR "fandom") AND ("Twitter" OR "X" OR "social media") AND ("Naive Bayes" OR "classifier" OR "machine learning" OR "text mining").

The search was not restricted to articles that explicitly used the phrase K-pop concert. Highly specific concert literature remains limited, so studies of groups, fandom, the Korean Wave, cyberbullying, and social-media sentiment were considered when they contributed an object, method, preprocessing procedure, evaluation strategy, or contextual explanation relevant to the review.

Priority was given to peer-reviewed national journals accredited in SINTA and international journals indexed in Scopus. Publications from 2022 to 2026 were emphasized to represent recent platform and methodological developments. Earlier conceptual work was retained only when it directly explained K-pop communities or provided a foundational account of sentiment analysis, information retrieval, or systematic-review reporting.

2.3. Inclusion and Exclusion Criteria

The selection criteria were designed to preserve topical relevance, source quality, methodological contribution, and recency. They also distinguish directly relevant concert studies from supporting studies that contribute platform, method, or fandom context. Table 1 summarizes the criteria.

Table 1. Article inclusion and exclusion criteria

Aspect	Inclusion	Exclusion
Topic	Sentiment analysis, opinion, emotion, K-pop, the Korean Wave, concerts, fandom, or text mining.	No relationship to sentiment analysis, K-pop/Korean Wave, social media, or text classification.
Media/data	X/Twitter, social media, or digital text data.	No digital text data or no relevance to social media.
Method	Naive Bayes, SVM, deep learning, transformers, NLP, or a relevant comparative method.	No contribution to sentiment analysis or the K-pop/X context.
Source quality	A SINTA-accredited national journal or a Scopus-indexed international journal.	Unclear source, no peer review, or unsuitable index status.
Period	Priority to publications from 2022–2026; highly relevant conceptual work was retained.	Outside the period without a direct conceptual contribution or with ambiguous metadata.

2.4. PRISMA Selection Process

The article selection process followed the PRISMA workflow. The identification phase yielded 5,674 records derived from 20 search export files (in CSV format), which were generated using various keyword combinations across the respective databases. These files were merged into a single initial dataset prior to deduplication. Deduplication was performed automatically using Mendeley reference management software, based on matching DOIs and normalized titles (case-insensitive and with punctuation removed). Following the automated process, researchers manually reviewed the results to ensure no duplicates were missed and no articles were erroneously removed. This process resulted in 1,056 unique records. This stage is crucial, as duplication can compromise research accuracy and lead to redundant processing of the same article. After deduplication based on DOIs and normalized titles, 4,618 records were eliminated, leaving 1,056 unique records.

During the title and abstract screening phase, 908 records were excluded for failing to meet criteria regarding the topic, publication type, media/data, or required methodology. This stage left 148 candidate articles. Subsequently, an eligibility assessment was conducted, evaluating articles based on SINTA accreditation or Scopus indexing, publication

year, availability of full text or metadata, and relevance to K-pop, the Korean Wave, concerts, fandoms, X/Twitter, sentiment analysis, and text classification.

Of the 148 candidate articles, 128 were eliminated during the eligibility phase. Consequently, a final set of 20 articles was selected for synthesis. This number was deemed sufficient for thematic mapping, as it encompasses national articles offering robust sentiment analysis methodologies alongside international articles that provide context regarding fandoms and social media. The detailed article-selection process is illustrated in Figure 1, while the summary of the PRISMA selection process is presented in Table 2.

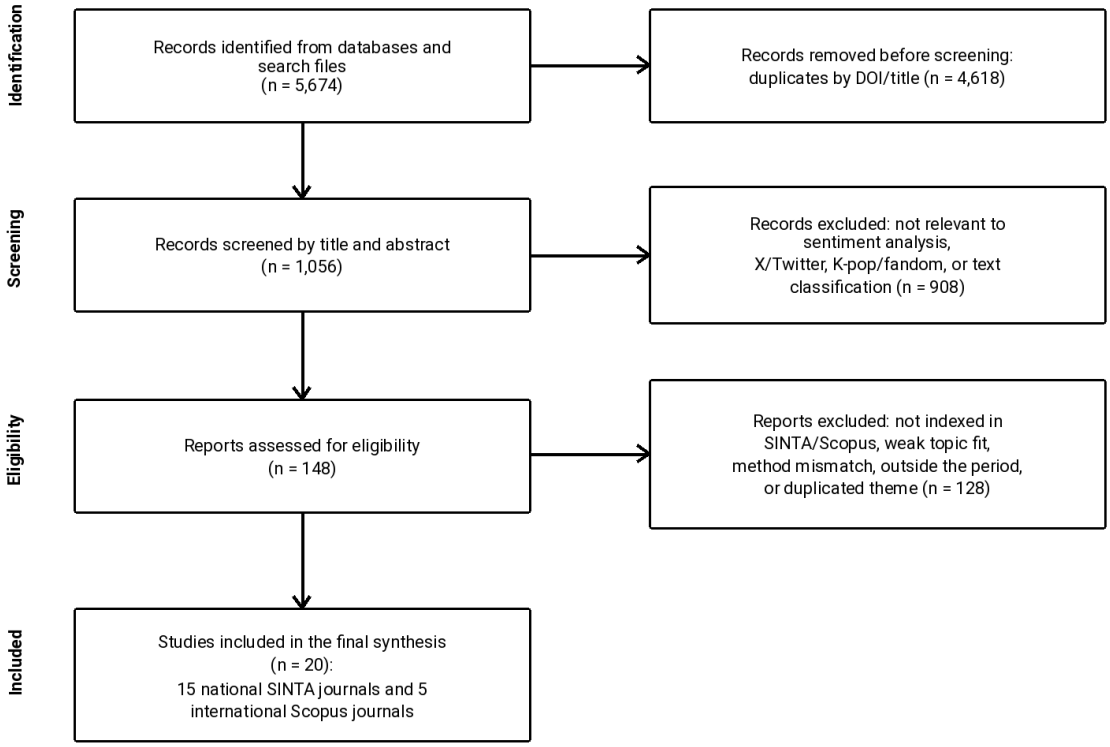

Figure 1. PRISMA diagram of the article-selection process

Table 2. Summary of the PRISMA selection process

Stage	Entered	Excluded	Remaining	Decision basis
Identification	5,674	0	5,674	All records from 20 search-result files were collected.
Deduplication	5,674	4,618	1,056	Duplicate DOI values or normalized titles were removed.
Title & abstract screening	1,056	908	148	Topic, media, publication type, and method relevance were checked.
Eligibility	148	128	20	Indexing, year, text/metadata access, and focus were verified.
Included	20	0	20	The final articles were synthesized narratively and thematically.

2.5. Data Extraction and Analysis

Data extraction uses a structured matrix containing author and year, title, journal, object, platform, algorithm, preprocessing, dataset, evaluation metrics, principal findings, contribution, and limitation. The matrix made comparison possible even when the original articles used different terminology and experimental procedures.

The extracted information was synthesized narratively and thematically. The first theme relates to developments in publication years, platforms, and research objects. The second compares algorithms and preprocessing procedures. The third examines performance measures and experimental reporting. The fourth identifies strengths, limitations, and research gaps that should guide concert-specific studies.

Relevance was interpreted as a continuum rather than a binary property. A study of a K-pop concert on X was considered directly relevant, whereas a study of cyberbullying on X was methodologically relevant, and a study of fandom participation was contextually relevant. This classification prevents peripheral evidence from being treated as equivalent to concert-focused evidence while still preserving useful methodological lessons.

3. Results and Discussion

This section presents and discusses the findings obtained from the systematic review of the selected studies. The discussion is organized around the general profile of the studies, research developments and study objects, analytical methods and model performance, as well as the strengths, limitations, and research gaps identified in the literature. Based on these findings, a recommended research design and the academic and practical implications of the study are also presented.

3.1. Profile of the Selected Studies

The final set of 20 articles consists of national and international research published mainly between 2023 and 2026. National studies contribute detailed applications of Naive Bayes, SVM, TF-IDF, and Indonesian-language preprocessing. International studies strengthen the conceptual understanding of fan communities, livestream concerts, social interaction, platformization, and transnational information diffusion.

X/Twitter is the dominant data source because it provides public, time-sensitive, text-based opinion. YouTube appears in one study of Blackpink and illustrates a different form of platform discourse in which comments remain attached to a specific video. The dominance of X is particularly relevant to concerts because event discussions change rapidly across ticket sales, event days, and post-event evaluations.

The methods range from classical machine learning to deep learning, transformers, qualitative research, and social-network analysis. Naive Bayes is used as a main classifier or comparator in several studies. SVM is the most frequent classical competitor, while CNN, IndoBERT, and XLM-RoBERTa represent the more recent movement toward contextual representation.

The selected studies cover various K-pop-related research objects. Some articles study Treasure, BTS, NCT, or Blackpink; others investigate Korean drama, the broader Korean Wave, cyberbullying, digital nationalism, fandom, or general policy opinion. This diversity provides a useful methodological base, but it also indicates that comprehensive studies focusing specifically on K-pop concert experiences remain limited.

The synthesis table should not be interpreted as a direct ranking of algorithms. Dataset size, annotation procedures, class distribution, preprocessing, train-test split, and evaluation design differ substantially. Reported accuracy is therefore used to describe each study, while comparative conclusions are limited to studies that evaluated models under the same experimental conditions. To provide a structured overview of the selected studies, including research objects and platforms, methods applied, and main findings of the 20 included articles, Tabel 3 presents the synthetis of the selected studies.

Table 3. Synthesis of the 20 selected articles

Study	Object & platform	Method	Main result
Chulyatunni'mah et al., 2025 [12]	Treasure; 2024 Karnaval Mandiri; X	Naive Bayes, crawling, TextBlob	1,018 tweets; neutral 38.1%, positive 35.3%, negative 26.6%; 96% accuracy.
Putri & Muzakir, 2022 [7]	K-pop cyberbullying; Twitter	Naive Bayes	Comments were classified into positive, negative, and neutral sentiment.
Savitri et al., 2024 [9]	Korean Wave in Indonesia; X	CNN / deep learning	717,998 tweets; 79% accuracy.
Riyadi et al., 2024 [10]	Blackpink; YouTube	IndoBERT	98% accuracy; F1, recall, and accuracy scores of 95%.
Aprilia & Lestari, 2024 [11]	Korean drama; X	Naive Bayes	88.44% precision; 91.70% positive recall; 84.33% accuracy.
Riyandona et al., 2025 [13]	BTS hiatus; Twitter	Naive Bayes, TF-IDF	78.33% accuracy; 79.25% precision; 78.33% recall; 78.49% F1; positive sentiment dominated.
Safitri et al., 2023 [17]	BTS; Twitter	SVM	Three data-split scenarios; the best results were obtained with 90:10 and 80:20 splits.

Study	Object & platform	Method	Main result
Noviana & Rasal, 2023 [8]	BTS; Twitter	Naive Bayes and SVM	Naive Bayes reached 79% and SVM 81% accuracy; positive sentiment dominated.
Damayanti et al., 2024 [20]	NCT; X	SVM, TF-IDF	380 data points; 82% accuracy.
Malik & Haidar, 2023 [2]	K-pop stan community; Twitter	Interviews, observation, coding	Twitter fandom operates as a community of practice.
Kim et al., 2022 [1]	K-pop and global behavior; SNS	Survey, regression, SEM	1,247 respondents; K-pop factors influenced national image and SNS citizenship behavior.
Upham et al., 2024 [15]	BTS fans during livestream concerts; Twitter	Activity and content analysis	Examined fan engagement across four concerts using a sample of 1,200 tweets.
Sari et al., 2025 [14]	Examines bilingual cyberbullying on X and is included only as a methodological comparison of Naive Bayes and SVM	Naive Bayes and SVM	Naive Bayes performed better, reaching 87% accuracy on Indonesian data.
Witarti et al., 2026 [21]	Indonesian K-pop fandom; social media and concerts	Netnography, text mining, clustering	Three clusters were characterized by retweet, voice, pride, culture, donation, and concert terms.
Astari et al., 2024 [18]	BTS/ARMY fanbase; Twitter	Virtual ethnography, FGD	ARMY constructs symbolic reality through content and active participation.
Arfan et al., 2024 [23]	Cyberbullying; X	Naive Bayes	86% accuracy.
Ikhsan et al., 2025 [24]	Examines public opinion on a VAT increase on X and is included only to show the effect of class balancing on Naive Bayes and SVM	Naive Bayes, SVM, balancing	SVM reached 98% before and 97% after balancing; Naive Bayes reached 97% before and 90% after balancing.
Nam et al., 2024 [19]	Transnational BTS ARMY; Twitter	Social network analysis	Identified key actors and information-diffusion patterns in the fandom system.
James, 2025 [16]	K-pop fandom on social-media platforms	Ten-month qualitative fieldwork	Developed the concept of affective participation in platformized fandom.
Gutiérrez-Jauregi et al., 2025 [22]	K-pop, social media, and fandom dynamics	Literature review and survey of 315 fans	Social media shapes emotional ties, fan behavior, narratives, and social impact.

3.2. Research Development and Study Objects

Research published in 2022 and 2023 relied primarily on classical machine-learning models. The principal objective was to demonstrate that tweets could be classified into positive, negative, and neutral categories after cleaning and feature extraction. Putri and Muzakir used Naive Bayes for K-pop cyberbullying, while Noviana and Rasal compared Naive Bayes with SVM for BTS-related Twitter data [7], [8].

In 2024, both research objects and methods became more diverse. Savitri et al. applied a CNN to a large X dataset concerning the influence of South Korean culture in Indonesia [9]. Riyadi et al. used IndoBERT for Blackpink-related YouTube comments [10], and Aprilia and Lestari applied Naive Bayes to Korean-drama discussion on X [11]. These studies extend the domain from music groups to the broader Korean Wave.

The 2025 literature contains work closer to the present topic. Chulyatunni'mah et al. analyzed fan sentiment toward Treasure at Karnaval Mandiri using Naive Bayes [12]. Riyandona et al. evaluated BTS discussion during a hiatus [13], and Sari et al. compared Naive Bayes with SVM for bilingual cyberbullying on X [14]. The continued use of Naive Bayes indicates that classical methods remain relevant even as contextual models become more common.

International studies add a social interpretation that classification research often lacks. Upham et al. examined BTS fan interaction during livestream concerts, James discussed the platformization of K-pop fandom, and Malik and Haidar described K-pop stan Twitter as a community of practice [2], [15], [16]. These studies show that posts are components of collective practices rather than isolated expressions.

The most concert-relevant objects are Treasure at Karnaval Mandiri, BTS livestream concerts, and NCT concert discussion. Jove and Paramita used XLM-RoBERTa to analyze NCT concert sentiment and identified information, ticket fraud, and technical constraints as important issues [5]. However, a study that combines a specific Indonesian concert, X data, and machine-learning-based sentiment analysis remains difficult to find.

BTS dominates several strands of the literature because of its large global fan base and sustained social-media activity [8], [13], [15], [17], [18], [19]. This concentration provides extensive evidence about one fandom but may limit generalization to other groups, local promoters, or smaller concert communities. Future datasets should therefore state clearly whether conclusions concern one artist, one event, or K-pop concerts more broadly.

Blackpink and NCT studies demonstrate that platform and object influence the form of available opinion. YouTube comments are anchored to a video and may accumulate over time, whereas X supports hashtags, threads, reposts, and live commentary. Event-focused research on X can therefore observe temporal changes that are less visible in a static comment section.

Studies of the Korean Wave and Korean drama contribute contextual evidence concerning public acceptance of Korean cultural products [9], [11]. Cyberbullying studies contribute methodological lessons about negative language, informal spelling, and class imbalance [7], [14], [23]. Fandom studies contribute the community-level meaning needed to interpret internal expressions, coordinated posting, and symbolic language [2], [16], [18], [19], [21], [22].

Overall, RQ1 shows a movement from classical classification of general objects toward a combination of contextual models and more specific cultural phenomena. The literature is rich in K-pop and social-media studies but uneven in concert coverage. A concert is analytically distinct because it combines transactions, physical access, service quality, crowd experience, performance evaluation, and digital community response.

3.2.1. Temporal and Publication Patterns

The publication pattern also reflects a change in the questions asked by researchers. Earlier studies primarily asked whether social-media text could be separated into polarity classes. Later work increasingly asked which model captured context more effectively, how fandom activity should be interpreted, and how digital participation was connected with broader cultural behavior. This transition indicates that the field is moving from proof-of-concept classification toward domain-sensitive and socially informed analysis.

The change from the name Twitter to X creates a minor but important reporting issue. Articles using older datasets retain the term Twitter, whereas more recent publications use X. Treating the terms as equivalent without recording the collection period may obscure platform changes, data-access policies, or user behavior. A future concert study should therefore report both the platform name used at collection time and the exact dates covered by the dataset.

Publication recency is useful, but methodological maturity cannot be inferred from year alone. A recent article may use a complex model while reporting limited annotation or validation details, and an older article may provide a more reproducible experimental pipeline. The synthesis therefore evaluates each study through its object, data, preprocessing, evaluation, and interpretive contribution rather than assuming that the newest algorithm is automatically the strongest evidence.

The national literature is particularly valuable for Indonesian preprocessing and practical implementation. It demonstrates how Naive Bayes, SVM, and TF-IDF can be applied to local-language data and provides performance benchmarks for student and institutional research. International literature complements this evidence by explaining fandom structures, livestream interaction, and transnational communication, areas that are often outside the scope of classification experiments.

The 20 studies should be understood as a map of adjacent evidence rather than a homogeneous experimental corpus. Only a subset directly addresses K-pop sentiment; an even smaller subset concerns concert. The inclusion of supporting studies is justified because a concert-focused method must still solve problems of multilingual text, class imbalance, cyberbullying vocabulary, community norms, and platform-specific communication.

A useful consequence of this mapping is the distinction between topic proximity and methodological relevance. The NCT concert study is topically close but uses a transformer, whereas a public-policy study is topically distant but demonstrates the impact of balancing on Naive Bayes [5], [24]. Both contribute to the proposed design, although they answer different parts of the methodological problem.

3.2.2. Platform and Object Coverage

Group-centered studies provide clear search terms and highly engaged communities, but they can also generate sampling bias. A dataset collected through an artist name or fandom hashtag may overrepresent committed fans and

underrepresent casual attendees, parents, venue workers, or users discussing a promoter without naming the artist. Query design should therefore combine artist, event, venue, promoter, and ticket-related terms.

Concerts differ from general K-pop discussion because the evaluated object is composite. A single post may praise the artist, criticize the sound system, report a ticket scam, and provide neutral schedule information. A one-label document classifier forces these components into one category. Aspect annotation or sentence-level segmentation can reduce this problem by identifying which component carries each sentiment.

Livestream concerts and physical concerts should not be treated as interchangeable. Livestream studies reveal online co-presence, platform interaction, and synchronized fan activity [15]. Physical events add travel, queues, security checks, seating, visibility, weather exposure, merchandise access, and crowd management. The proposed research scope should state clearly which event format is analyzed and which aspects are relevant.

Fandom research also warns that posting volume does not equal the number of independent opinions. Coordinated campaigns, repeated templates, fan projects, and repost networks can amplify particular phrases. Deduplication based only on identical text may not remove near-duplicate campaign messages. Researchers should decide whether coordinated participation is noise, an object of analysis, or a separate feature of the event discourse.

Cyberbullying literature is relevant because negative expressions may contain insults, coded language, or target-specific vocabulary. However, a concert complaint is not necessarily abusive, and an enthusiastic fan rivalry message may not evaluate the concert at all. Annotation guidelines should distinguish service criticism, interpersonal hostility, artist-directed attacks, and ordinary negative experience.

The broader Korean Wave literature contributes a cultural frame but should not substitute for event evidence. Positive attitudes toward Korean culture may influence prior expectations, yet satisfaction with a concert depends on operational delivery. A concert dataset is therefore needed to test whether general cultural enthusiasm remains positive when users discuss prices, access, facilities, or promoter performance. The classification of research objects, along with their contributions and limitations, is presented in Table 4 as follows.

Table 4. Research-object groups and their contributions

Group	Examples	Contribution and limitation
Specific groups/artists	Treasure, BTS, NCT, and Blackpink	Closest to fan responses, but often centered on major groups and rarely covers the complete concert experience.
General Korean Wave	South Korean culture and Korean drama	Explains acceptance of Korean cultural products, although its concert relevance is mainly contextual.
Social issues	Cyberbullying and public-policy opinion on X	Provides lessons about emotional language, class imbalance, and classification evaluation.
Digital fandom	Community, digital nationalism, networks, and affective participation	Explains symbols, norms, coordination, and internal vocabulary that shape sentiment interpretation.

3.3. Methods, Preprocessing, and Model Performance

The reviewed methods can be organized into four groups: classical machine learning, deep learning, transformers, and qualitative or social-computational approaches. Naive Bayes and SVM form the classical group; CNN represents deep learning; IndoBERT and XLM-RoBERTa represent contextual transformers; and netnography, virtual ethnography, content analysis, and social-network analysis provide interpretive context.

Naive Bayes is the most frequently recurring classical method in the reviewed literature. It appears in studies of Treasure, K-pop cyberbullying, Korean drama, BTS, bilingual cyberbullying, general cyberbullying, and public-policy opinion [7], [8], [11] - [14], [23], [24]. Its repeated use reflects low computational cost, straightforward implementation, and compatibility with sparse term-frequency or TF-IDF features.

The principal advantage of Naive Bayes is transparent probabilistic classification. Researchers can explain how term distributions contribute to class prediction, reproduce the pipeline with modest resources, and establish a baseline before applying more complex models. Its limitation is that the conditional-independence assumption simplifies relationships among words and does not directly represent long-range context, sarcasm, or compositional meaning.

SVM is the strongest recurring classical comparator. It is effective in high-dimensional sparse spaces and may form a clear decision boundary when TF-IDF features separate classes well. Noviana and Rasal reported 81% accuracy

for SVM and 79% for Naive Bayes [8]. In contrast, Sari et al. reported that Naive Bayes performed better on Indonesian bilingual-cyberbullying data [14]. The difference confirms that model superiority is dataset-dependent.

CNN learns local patterns without relying solely on manually designed term weights. Savitri et al. applied it to 717,998 tweets and reported 79% accuracy [9]. Although the result is lower than several smaller studies, the object, label distribution, and experimental difficulty differ. Deep-learning performance should therefore not be judged from accuracy alone or compared across unrelated datasets.

Transformer models provide contextual representations in which the meaning of a token depends on surrounding text. IndoBERT produced strong results for Blackpink comments, and XLM-RoBERTa was used for multilingual NCT concert discussion [5], [10]. These models are valuable for mixed languages and context-sensitive expressions, but they require more computation, careful fine-tuning, and representative training data.

Qualitative and network approaches do not produce automated sentiment labels, yet they explain why classification errors occur. Fandom members use symbolic references, coordinated hashtags, playful spelling, and in-group meanings that cannot be interpreted reliably through a general dictionary [2], [16], [18], [19], [21], [22]. These studies can inform annotation guidelines and domain-specific normalization.

3.3.1. Feature Representation and Baseline Suitability

For Multinomial Naive Bayes, the feature space is usually derived from term counts or TF-IDF values. The model estimates class-conditional term probabilities and combines them with prior class probabilities. Laplace or additive smoothing is important because concert-specific terms may appear in one class but not another. The smoothing value, vectorizer settings, minimum document frequency, and n-gram range should be reported because they directly influence predictions.

Unigrams provide a transparent baseline, but bigrams can preserve expressions such as “not safe,” “sold out,” “worth it,” or “bad sound.” Negation handling is especially important because removing a stop word such as “not” can reverse polarity. A controlled experiment can compare unigram and unigram-bigram features while keeping the classifier and data split unchanged.

Class priors should be examined rather than accepted automatically. Concert datasets may contain a large neutral class because many posts share schedules, ticket links, or venue directions. A model trained with empirical priors may favor neutral predictions. Researchers can compare empirical priors, balanced training, and class-specific sampling, but every intervention must be evaluated on an untouched test set.

Naive Bayes is also useful for interpretation. Terms with high log-probability differences between classes can reveal which expressions drive positive or negative predictions. These terms should be reviewed by domain-informed annotators because a fandom phrase may appear negative in ordinary language but positive within the community. Interpretation therefore connects statistical output with qualitative knowledge.

SVM comparison is methodologically valuable because both models operate effectively with sparse TF-IDF features but use different decision principles. When SVM outperforms Naive Bayes, error analysis can determine whether the gain occurs in boundary cases, minority classes, or longer posts. When Naive Bayes performs similarly, its lower complexity and easier explanation may justify its use.

Transformer comparison answers a different question: whether contextual representation resolves cases that bag-of-words models cannot. Examples include sarcasm, implicit sentiment, multilingual code-switching, and words whose meaning depends on an event reference. The comparison should focus not only on overall accuracy but on the specific error categories improved by contextual modeling.

Computational efficiency should be reported alongside predictive performance. Training time, inference time, memory requirements, and hardware affect whether a method can be reproduced by other researchers or used for rapid event monitoring. A small performance gain may not justify a much larger computational burden when the operational objective is transparent and timely analysis. To summarize the strengths, limitations, and recommended roles of the research approaches discussed above, a comparison is presented in Table 5, as follows.

Table 5. Comparison of research approaches

Approach	Strength	Limitation	Recommended role
Naive Bayes	Probabilistic, fast, reproducible, and suitable for high-dimensional text features.	Sensitive to independence assumptions, label quality, and class imbalance.	Primary baseline with TF-IDF and domain-aware preprocessing.

Approach	Strength	Limitation	Recommended role
SVM	Strong in sparse feature spaces and can outperform Naive Bayes on some datasets.	Requires parameter or kernel selection and is less interpretable.	Classical comparative model.
CNN / deep learning	Learns local patterns and feature representations automatically.	Requires more data and computation; results are hard to compare without a clear pipeline.	Comparator for large datasets.
IndoBERT / XLM-RoBERTa	Captures context, multilingual expressions, and word relationships more effectively.	High computational cost and requires fine-tuning with representative data.	Contextual comparator and error-analysis model.
Qualitative / network	Explains fandom norms, symbols, key actors, and participation context.	Does not provide automated sentiment classification.	Supports domain interpretation and annotation guidelines.

Preprocessing is a decisive component of model performance. Common stages include URL and mention removal, case folding, tokenization, normalization, stop-word removal, stemming, and TF-IDF weighting. However, reporting is inconsistent. Several articles mention preprocessing without providing the normalization dictionary, emoji rules, hashtag policy, or the sequence in which operations were applied.

Concert data requires selective rather than aggressive cleaning. An artist name, member name, venue, promoter account, ticket category, or hashtag may identify the aspect being evaluated. Emoji can express polarity, while repeated letters and capitalization may encode intensity. Removing these features automatically can reduce noise but also eliminate information that distinguishes a complaint from an enthusiastic reaction.

Normalization is particularly difficult because posts may combine Indonesian colloquial forms, English phrases, Korean romanization, and fandom-specific abbreviations. A useful dictionary must be developed from the target event and validated manually. Negation should be preserved, and stemming should not alter proper names or community terminology. These domain decisions should be reported so that results are reproducible.

Evaluation must extend beyond accuracy because social-media classes are commonly imbalanced. Precision, recall, F1-score, confusion matrices, and per-class results reveal whether a model ignores minority classes. Ikhsan et al. showed that balancing changed Naive Bayes performance from 97% to 90% in a public-policy dataset [24]. This result illustrates why class distribution and the effect of balancing must be examined explicitly.

3.3.2. Domain-Aware Preprocessing

Data collection is part of preprocessing because the query determines which language and aspects enter the corpus. Queries based only on a group name may retrieve unrelated fan content, while overly narrow event hashtags may miss complaints that mention only the promoter or venue. A pilot collection can identify additional aliases, spelling variants, and ticket-related terms before the final period begins.

URL removal is generally appropriate, but the domain of a link may provide information about ticketing, resale, news, or fraud. Researchers may remove the full URL while retaining a categorical feature such as official ticketing site, news source, or unknown domain. The same principle applies to mentions: the account string can be removed from lexical text but mapped to an aspect such as artist, promoter, venue, or customer service.

Emoji should be processed systematically. A simple strategy converts emoji to textual descriptions before vectorization, while a richer strategy uses emoji categories or sentiment scores as additional features. The decision must be consistent because emoji can strengthen, soften, or contradict the surrounding words. Removing all emoji would discard a common component of fandom expression.

Hashtags have at least three functions: event identification, topic labeling, and emotional or campaign expression. Segmenting a hashtag into words can improve lexical representation, but some official tags and fandom slogans should also be retained as complete tokens. The pipeline should document whether the hash symbol is removed, whether camel case is segmented, and how multilingual tags are handled.

Repeated characters and capitalization can be normalized while preserving an intensity feature. For example, a long sequence of vowels may be reduced to a standard form, but the repetition count can indicate excitement or frustration. Similarly, all-uppercase text may be lowercased for vocabulary consistency while a binary capitalization flag is retained.

Stemming is not always beneficial in this domain. Indonesian affixes can be reduced to improve feature consolidation, but aggressive stemming may damage English words, Korean romanization, names, and branded terms.

A protected-term list and language-aware processing can reduce such errors. An ablation study should test whether stemming improves validation performance rather than assuming it is necessary.

A protected-term list should be created through a documented three-step rule: (1) compile candidate terms from official artist, member, venue, promoter, event, ticket-category, and fandom glossaries; (2) retain a term when stemming or normalization changes its identity or aspect meaning in a pilot sample; and (3) validate each addition through agreement between at least two domain-informed annotators. Every entry should record the original form, aliases, language, semantic category, and the transformation that must be blocked. The list should be versioned, updated only after adjudication, and documented with the preprocessing protocol so that the rule can be reproduced.

Quoted posts, replies, and reposts require explicit treatment. Including quoted text can duplicate the original opinion, whereas excluding it may remove the context needed to interpret a response. The dataset can store original and user-added text separately, allowing researchers to classify the new comment while retaining the quoted content as contextual metadata. To summarize the general purpose and domain-specific considerations of each preprocessing stage for K-pop concert data, the details are presented in Table 6, as follows.

Table 6. Preprocessing implications for K-pop concert data

Stage	General purpose	Domain consideration
Cleaning	Removes URLs, mentions, repeated characters, and non-informative technical elements.	Hashtags and promoter accounts should not be removed automatically because they may encode important aspects.
Case folding	Standardizes letter case so capitalization variants are not treated as different features.	Extreme capitalization may signal emotion and can be retained as an auxiliary feature.
Tokenization	Splits text into word or subword units.	The process should handle emoji, hashtags, compounds, and multilingual expressions.
Normalization	Maps slang, abbreviations, and non-standard spelling to canonical forms.	The dictionary should cover Indonesian, English, Korean, fandom vocabulary, and concert terminology.
Stop-word removal	Reduces function words considered minimally informative.	It must be selective so that negation and intensifiers are preserved.
Stemming	Reduces inflected words to their roots.	Artist, venue, promoter, and fandom terms should be protected from incorrect stemming.
TF-IDF / features	Weights terms that distinguish documents and sentiment classes.	It can be extended with n-grams, emoji, hashtags, event phases, and concert aspects.

The reported performance range is wide. Naive Bayes achieved 96% in the Treasure study, 84.33% in the Korean-drama study, and 78.33% in the BTS-hiatus study [11] - [13]. The variation should not be interpreted as instability of the algorithm alone. Dataset size, annotation, topic complexity, language variation, class balance, and split strategy can all produce substantial differences.

A strong future experiment should use Multinomial Naive Bayes with TF-IDF as a reproducible baseline, SVM as a classical comparator, and a transformer as a contextual comparator when resources allow. The same labeled dataset, train-test partitions, and evaluation measures must be used for every model. This design reveals whether added complexity produces meaningful improvement on concert discourse.

3.3.3. Evaluation and Error Analysis

A reliable evaluation begins with a split that prevents leakage. Near-duplicate posts, campaign templates, or repeated quotations should not be divided between training and test sets because the model may memorize wording. When timestamps are available, a chronological test can provide a more realistic estimate of performance on later event phases.

Macro-averaged F1 is particularly informative when positive, negative, and neutral classes differ in size because it gives equal importance to each class. Weighted F1 reflects overall distribution but may conceal poor minority performance. Reporting both measures, together with per-class precision and recall, allows readers to understand the trade-off.

Confusion matrices should be interpreted rather than displayed without discussion. Confusion between neutral and positive may arise from informational posts containing enthusiastic emoji, while confusion between negative and

neutral may arise from implicit complaints. Misclassification between positive and negative is more serious and often indicates sarcasm, negation failure, or mixed-aspect posts.

Cross-validation can reduce dependence on one random split, but it must respect duplicate groups and temporal structure. Stratified folds preserve class proportions, whereas group-based folds can keep near-duplicate campaigns together. A final held-out test set should remain untouched if model settings are selected through repeated validation.

Statistical significance and confidence intervals are rarely reported in the reviewed literature. For closely performing models, bootstrap intervals or paired tests can indicate whether a difference is stable or may result from sampling variation. This is important when deciding whether a complex transformer meaningfully improves on a transparent baseline.

Error analysis should produce a taxonomy that includes sarcasm, code-switching, domain terminology, implicit sentiment, mixed aspects, annotation disagreement, and insufficient context. Reporting representative anonymized examples for each category provides more methodological value than presenting accuracy alone and directly informs future preprocessing and annotation revisions. To provide a concise comparison of the key quantitative findings reported in the selected studies, Tabel 7 summarizes the main objects, models, results, and relevant notes.

Table 7. Summary of key quantitative results

Object	Model	Result	Note
Treasure/Karnaval Mandiri	Naive Bayes	96% accuracy	Neutral 38.1%, positive 35.3%, and negative 26.6% [12].
Influence of Korean culture	CNN	79% accuracy	A very large dataset was used, but the object represented the broader Korean Wave [9].
Blackpink/YouTube	IndoBERT	98% accuracy	F1, recall, and accuracy scores were reported at approximately 95% [10].
Korean drama/X	Naive Bayes	84.33% accuracy	Precision reached 88.44% and positive-class recall reached 91.70% [11].
BTS hiatus	Naive Bayes + TF-IDF	78.33% accuracy	Precision 79.25%, recall 78.33%, and F1 78.49% [13].
BTS/Twitter	Naive Bayes vs SVM	79% vs 81%	SVM was slightly higher on this dataset [8].
NCT/X	SVM + TF-IDF	82% accuracy	The evaluation used 380 data points [20].
Bilingual cyberbullying	Naive Bayes vs SVM	NB 87%	Naive Bayes performed better on the Indonesian-language data [14].

3.4. Strengths, Limitations, and Research Gaps

The literature has several strengths. It covers diverse artists, platforms, cultural objects, and analytical approaches. National studies provide practical procedures for Indonesian text classification, while international studies explain fandom participation and community structure. Together, they establish that X is both a computational data source and a social space with distinctive norms.

The first limitation is object specificity. Most studies focus on a group, fandom, cultural trend, or social issue rather than the complete concert experience. Tickets, promoter communication, venue facilities, safety, queues, technical production, artist performance, and post-event service are usually examined separately or not at all. This prevents a comprehensive account of what produces positive or negative concert sentiment.

The second limitation concerns data and annotation. Articles do not always describe collection dates, query terms, duplicate removal, bot filtering, manual-labeling guidelines, annotator qualifications, or inter-annotator agreement. These omissions make it difficult to determine whether reported performance reflects a robust classifier or an unusually simple and homogeneous dataset.

The third limitation concerns preprocessing transparency. Fandom slang, multilingual terms, emoji, hashtags, and sarcasm are central to the domain, but dictionaries and transformation rules are rarely published. A pipeline designed for general Indonesian text may incorrectly normalize artist names, remove meaningful negation, or discard the hashtag that identifies the event and aspect.

The fourth limitation is evaluation comparability. Accuracy values from different studies cannot be ranked directly because datasets, labels, class balance, and validation procedures differ. Some studies report precision, recall, and F1-

score, whereas others provide accuracy alone. Few presents per-class confusion matrices or detailed error analysis, even though minority negative sentiment may be the most operationally important class.

The fifth limitation is interpretation. A negative post about ticket fraud or crowd management does not necessarily express a negative attitude toward the artist. Conversely, a positive post praising a performance does not imply satisfaction with the promoter. Polarity labels should therefore be combined with concert aspects so that organizational criticism is separated from artist-related sentiment.

The central research gap is a study that directly combines an Indonesian K-pop concert, temporally bounded X data, domain-aware preprocessing, Naive Bayes classification, validated annotation, per-class evaluation, and aspect-based interpretation. The Treasure study is event-related and reports strong accuracy, but it does not represent the full concert journey [12]. The NCT concert study is more specific, but it uses XLM-RoBERTa rather than Naive Bayes [5].

Future data collection should be divided into pre-concert, during-concert, and post-concert phases. The pre-concert phase captures ticketing, schedules, promoter communication, and preparation. The event phase captures queues, venue access, safety, sound, screens, and artist performance. The post-concert phase captures satisfaction, complaints, refunds, lost property, and willingness to attend future events.

Annotation should combine polarity with aspects such as ticketing, promoter, venue, safety, performance, facilities, and audience experience. At least two annotators should apply written guidelines, and agreement should be reported. Sarcasm, mixed languages, quoted posts, reposts, and ambiguous informational messages require explicit rules. A small adjudicated gold set can support consistent training and evaluation.

Model evaluation should report precision, recall, F1-score, macro and weighted averages, confusion matrices, and class-specific errors. Experiments should document the train-test split, random seed, feature configuration, and balancing strategy. Error analysis should identify whether mistakes are caused by sarcasm, multilingual text, implicit sentiment, domain terminology, or disagreement among annotators.

With this design, sentiment analysis becomes more than a count of positive and negative posts. It can identify operational sources of satisfaction and complaint, compare event phases, and provide evidence for promoters, venue managers, fan communities, and researchers. The combination of a transparent baseline and contextual interpretation also makes conclusions more defensible and practically useful. To provide an overall assessment of the reviewed literature, the main strengths and limitations related to research objects, data and annotation, preprocessing, models, and evaluation are summarized in Table 8, as follows.

Table 8. Strengths and limitations of the literature

Aspect	Strength	Limitation
Object and context	The literature covers diverse K-pop, fandom, and social-media objects.	Concerts as complete experiences—tickets, promoters, venues, safety, and performance—remain rarely analyzed.
Data and annotation	X provides timely and opinion-rich data.	Sample size, labeling procedures, inter-annotator agreement, and class distribution are often incompletely reported.
Preprocessing	Several studies apply systematic text-processing pipelines.	Normalization dictionaries, emoji/hashtag rules, and protection of fandom terminology are not consistently documented.
Models	Naive Bayes, SVM, CNN, and transformers provide varied alternatives.	Comparisons often use different datasets and scenarios, preventing direct performance comparison.
Evaluation	Many studies report accuracy and some additional metrics.	Per-class results, confusion matrices, cross-validation, and error analysis remain limited.
Interpretation	Social studies explain fandom dynamics.	Quantitative classification often fails to separate sentiment toward artists from criticism of event organization.

3.5. Recommended Research Design

Based on the synthesis, future research should combine phase-based concert data, polarity-and-aspect annotation, domain-aware preprocessing, model comparison on the same dataset, per-class evaluation, and error analysis. To provide a clear framework for future studies, the recommended research design covering data collection, annotation, preprocessing, models, evaluation, and interpretation is summarized in Table 9, as follows.

Table 9. Recommended design for future research

Component	Recommendation
Data	Collect tweets from pre-concert, during-concert, and post-concert phases of a specific Indonesian event.
Annotation	Assign positive, negative, and neutral labels together with ticket, promoter, venue, safety, performance, and experience aspects.
Preprocessing	Normalize multilingual text and fandom slang; retain informative hashtags/emoji; apply selective stop-word removal and stemming.
Models	Use Multinomial Naive Bayes with TF-IDF as the baseline and compare it with SVM and transformers.
Evaluation	Report precision, recall, F1-score, confusion matrices, per-class results, and the impact of balancing.
Interpretation	Analyze influential terms and model errors; separate sentiment toward artists from event organization.

3.6. Academic and Practical Implications

The synthesis has an academic implication: concert sentiment analysis should be treated as a domain-specific classification problem rather than a generic application of a standard pipeline. The event lifecycle, fandom vocabulary, and separation between artist and organizer sentiment affect data collection, annotation, preprocessing, evaluation, and interpretation. A study that ignores these relationships may produce technically correct labels but misleading conclusions.

The proposed framework is modular. Data collection defines the event and phases; annotation assigns polarity and aspects; preprocessing converts multilingual social text into reproducible features; models provide baseline and contextual comparisons; evaluation identifies class-specific strengths; and interpretation links model output to concert operations. Because each component is explicit, its contribution can be tested through controlled ablation or sensitivity analysis.

For promoters, aspect-level results can identify recurring operational concerns. Ticketing, queue management, venue access, facilities, security, production quality, and customer service can be evaluated separately from artist performance. Temporal analysis can show whether a problem begins during ticket sales, emerges at the venue, or continues after the event through refund or complaint processes.

For fan communities, transparent analysis reduces the risk that coordinated expression or internal language is misrepresented. Domain-informed annotation can distinguish fan projects, jokes, affectionate nicknames, rivalry language, and genuine hostility. Publishing clear coding rules also allows community members and researchers to understand how posts were transformed into analytical categories.

For model developers, the dataset offers a realistic test of multilingual and informal language. Comparing Naive Bayes, SVM, and a transformer on identical data can reveal which error types require contextual modeling and which can be solved through better normalization or aspect design. This evidence is more useful than assuming that the most complex architecture is necessarily best.

For future systematic reviews, standardized reporting would improve comparability. Articles should provide collection dates, queries, dataset size after each cleaning stage, class distribution, annotation agreement, preprocessing settings, model parameters, train-test procedures, all major metrics, and representative errors. Consistent reporting would allow later reviews to conduct stronger quantitative comparisons or meta-analysis.

The practical value of the research depends on ethical handling of public posts. Usernames and identifying details should be removed from examples, data storage should be controlled, and quotations should be minimized or paraphrased when publication could make a user searchable. Ethical reporting is compatible with reproducibility when procedures and aggregate results are documented without exposing individual users.

Ethical assessment should address contextual integrity rather than assuming that public accessibility automatically removes risk. Researchers should collect only posts needed for the stated questions, comply with platform terms and applicable institutional requirements, avoid reporting handles or searchable verbatim quotations, and assess heightened risks for minors, marginalized users, and coordinated fan communities. When examples are necessary, they should be paraphrased and stripped of timestamps, locations, profile details, and other metadata that could enable re-identification.

The protocol should also document data minimization, secure storage, access control, retention and deletion schedules, treatment of deleted or private posts, and procedures for responding to withdrawal requests where feasible. Aspect-level findings should be reported in aggregate to avoid profiling individual users or stigmatizing fandoms. Any

shared dataset should contain only content that is authorized for redistribution or has been transformed sufficiently to protect users while preserving analytical value.

Ultimately, the recommended design connects methodological transparency with stakeholder relevance. It preserves Naive Bayes as an interpretable baseline, uses stronger models when they provide demonstrated benefits, and treats fandom knowledge as part of valid analysis rather than an informal addition. This combination can produce findings that are replicable, contextually accurate, and useful for improving future K-pop concert experiences.

4. Conclusion

This systematic review synthesized 20 articles concerning K-pop sentiment, the Korean Wave, fandom, cyberbullying, and social-media opinion. X/Twitter was the dominant platform, and publications increasingly combined classical classifiers with deep-learning and transformer models. However, the literature remained more focused on artists, communities, or broad cultural issues than on concerts as complete service and audience experiences.

Naive Bayes remains a widely used baseline in the reviewed studies because it is efficient, reproducible, and suitable for high-dimensional text. Reported performance ranged widely, demonstrating that accuracy depends on data quality, annotation, class balance, preprocessing, and feature representation. SVM and transformer models may provide stronger results under particular conditions, but valid comparison requires the same dataset and evaluation design.

Preprocessing should be domain-aware. Multilingual expressions, slang, emoji, hashtags, artist names, venue names, and promoter references can carry both aspect and polarity information. Evaluation should therefore include per-class precision, recall, F1-score, confusion matrices, and error analysis rather than accuracy alone.

The clearest research gap is an Indonesian concert study that combines X data from multiple event phases, validated polarity-and-aspect annotation, Multinomial Naive Bayes with TF-IDF, comparison with SVM and a transformer, and interpretation that separates artist sentiment from organizational criticism. Such a study can contribute methodologically while also producing actionable findings for event stakeholders.

Future research should report its query strategy, preprocessing rules, annotation guidelines, class distribution, parameter settings, and evaluation results in sufficient detail for replication. By combining computational transparency with knowledge of fandom culture, sentiment analysis can provide a more accurate account of the digital experience surrounding K-pop concerts.

References

[1] J. H. Kim, K. J. Kim, B. T. Park, and H. J. Choi, "The Phenomenon and Development of K-Pop: The Relationship between Success Factors of K-Pop and the National Image, Social Network Service Citizenship Behavior, and Tourist Behavioral Intention," Sustainability, vol. 14, no. 6, Mar. 2022, doi: https://doi.org/10.3390/su14063200.

[2] Z. Malik and S. Haidar, "Online community development through social interaction—K-Pop stan twitter as a community of practice," Interactive Learning Environments, vol. 31, no. 2, pp. 733–751, 2023, doi: https://doi.org/10.1080/10494820.2020.1805773.

[3] B. Liu, Sentiment Analysis and Opinion Mining. San Rafael, CA, USA: Morgan & Claypool, 2012, doi: https://doi.org/10.2200/S00416ED1V01Y201204HLT016.

[4] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge, U.K.: Cambridge University Press, 2008, doi: https://doi.org/10.1017/CBO9780511809071.

[5] C. E. Jove and A. S. Paramita, "Sentiment Analysis of K-pop Fans Toward NCT Concerts on X (Twitter) Using the Transformer Model XLM-RoBERTa," Journal of Applied Informatics and Computing, vol. 10, no. 2, pp. 1799–1805, Apr. 2026, doi: https://doi.org/10.30871/jaic.v10i2.12333.

[6] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, p. n71, Mar. 2021, doi: https://doi.org/10.1136/bmj.n71.

[7] A. Putri and A. Muzakir, "Analisis Sentimen Cyberbullying KPOP di Media Sosial Twitter Menggunakan Metode Naive Bayes," Syntax Literate: Jurnal Ilmiah Indonesia, vol. 7, no. 9, pp. 12421–12432, Sep. 2022, doi: https://doi.org/10.36418/syntax-literate.v7i9.9334.

[8] R. Noviana and I. Rasal, "Penerapan Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Boy Band BTS pada Media Sosial Twitter," Jurnal Teknik dan Science, vol. 2, no. 2, pp. 51–60, Jun. 2023, doi: https://doi.org/10.56127/jts.v2i2.791.

[9] P. Savitri, I. M. A. D. Suarjaya, and I. Vihikan, "Sentiment Analysis of X (Twitter) Comments on The Influence of South Korean Culture in Indonesia," Journal of Information Systems and Informatics, vol. 6, no. 2, pp. 979–991, Jun. 2024, doi: https://doi.org/10.51519/journalisi.v6i2.749.

[10] S. Riyadi, N. Salsabila, F. P. Damarjati, and A. Abdul Karim, "Sentiment Analysis of YouTube Users on Blackpink Kpop Group Using IndoBERT," INTENSIF, vol. 8, no. 2, pp. 233–245, Aug. 2024, doi: https://doi.org/10.29407/intensif.v8i2.22678.

[11] A. Aprilia and W. Lestari, "Analisa Sentimen Drama Korea Melalui Media Sosial X dengan Menggunakan Algoritma Naive Bayes," Jurnal Indonesia: Manajemen Informatika dan Komunikasi, vol. 5, no. 3, pp. 3248–3261, Sep. 2024, doi: https://doi.org/10.35870/jimik.v5i3.997.

[12] Chulyatunni'mah, R. Kurniawan, and S. Anwar, "Analisis Sentimen Penggemar Treasure di Karnaval Mandiri Menggunakan Naive Bayes," Jurnal Sistem Informasi Triguna Dharma, vol. 4, no. 1, pp. 203–213, Jan. 2025, doi: https://doi.org/10.53513/jursi.v4i1.10606.

[13] Riyandona, Rahaningsih, R. D. Dana, and Mulyawan, "Implementasi Model Analisis Sentimen Terhadap Grup Musik BTS Menggunakan Metode Naive Bayes," Jurnal Informatika dan Teknik Elektro Terapan, vol. 13, no. 1, pp. 1036–1050, Jan. 2025, doi: https://doi.org/10.23960/jitet.v13i1.5816.

[14] R. Sari, M. Jazman, T. K. Ahsyar, Syaifullah, and A. Marsal, "Penerapan Algoritma Klasifikasi Naive Bayes dan Support Vector Machine untuk Analisis Sentimen Cyberbullying Bilingual di Aplikasi X," Sistemasi, vol. 14, no. 1, pp. 211–224, 2025, doi: https://doi.org/10.32520/stmsi.v14i1.4799.

[15] L. Upham, Y. Lee, and S. Park, "Audience reconstructed: Social media interaction by BTS fans during live stream concerts," Frontiers in Psychology, vol. 15, Art. no. 1214930, Apr. 2024, doi: https://doi.org/10.3389/fpsyg.2024.1214930.

[16] K. James, "Affective Participation from the In-Between: The Platformization of K-Pop Fandom," Social Media + Society, vol. 11, no. 2, Jun. 2025, doi: https://doi.org/10.1177/20563051251351390.

[17] S. D. Safitri, Y. N. Umaidah, and I. Maulana, "Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine," Journal of Applied Informatics and Computing, vol. 7, no. 1, pp. 34–41, Jul. 2023, doi: https://doi.org/10.30871/jaic.v7i1.5039.

[18] N. Astari, N. Agustina, and E. Nurussa'adah, "Symbolic Reality Construction of the K-Pop Community on Twitter," Interaksi, vol. 13, no. 1, pp. 152–168, Jun. 2024, doi: https://doi.org/10.14710/interaksi.13.1.152-168.

[19] K. Nam, H. Kim, S. Kang, and H.-J. Kim, "The BTS ARMY on Twitter flocks together: How transnational fandom on social media build a viable system," Telematics and Informatics, 2024, doi: https://doi.org/10.1016/j.tele.2024.102143.

[20] E. Damayanti, Junaedy, and B. Herlinah, "Analisis Sentimen Penggemar Grup K-Pop NCT pada Media Sosial X (Twitter) Menggunakan Algoritma Support Vector Machine," Jurnal Teknologi dan Komputer, vol. 4, no. 2, pp. 483–490, Dec. 2024, doi: https://doi.org/10.56923/jtek.v4i02.224.

[21] Witarti, A. F. Putri, and E. Ariyani, "Aktivitas Nasionalisme Digital: Studi Netnografi Fandom K-Pop di Indonesia," Komuniti, vol. 18, no. 1, pp. 1–28, Mar. 2026, doi: https://doi.org/10.23917/komuniti.v18i1.13303.

[22] A. Gutiérrez-Jauregi, M. E. Aramendia-Muneta, and I. Gómez-Cámara, "Harmony in diversity: Unraveling the global impact of K-Pop through social media and fandom dynamics," Media Asia, pp. 1–27, Apr. 2025, doi: https://doi.org/10.1080/01296612.2025.2480451.

[23] A. R. Arfan, F. Fauziah, and I. Nawangsih, "Analisa Sentimen Terhadap Cyber Bullying di X Menggunakan Algoritma Naive Bayes," MALCOM, vol. 4, no. 4, pp. 1411–1419, Oct. 2024, doi: https://doi.org/10.57152/malcom.v4i4.1550.

[24] M. A. Ikhsan, N. P. Subarkah, Y. Iftinani, and A. N. Fadilah, "Analisis Sentimen Kenaikan PPN Menggunakan Algoritma Naive Bayes dan Support Vector Machine," Infotekmesin, vol. 16, no. 1, pp. 181–189, Jan. 2025, doi: https://doi.org/10.35970/infotekmesin.v16i1.2518.