Sentiment Analysis Related to Korean Subculture in Indonesia Using BERT Method

Sri Muspani[a1], I Made Agus Dwi Suarjaya[a2], Ni Kadek Dwi Rusjayanthi[a3]
[a]Department of Information Technology, Faculty of Engineering, Udayana University
Bukit Jimbaran, Bali, Indonesia-8036110
e-mail: [1]sri.muspani154@student.unud.ac.id, [2] agussuarjaya@it.unud.ac.id,
[3]dwi.rusjayanthi@unud.ac.id

Abstrak

Korean subculture mengalami peningkatan pesat di Indonesia dalam beberapa tahun terakhir. Masyarakat Indonesia memberikan pandangan positif, negatif maupun netral melalui media sosial X (Twitter). Analisis sentimen dilakukan untuk mengidentifikasi pandangan masyakarat sebagai positif, negatif atau netral. Tahapan analisis mencakup pengumpulan data, pre-processing data, pelabelan data, pelatihan model, pengklasifikasian, dan visualisasi. Data yang dianalisis berjumlah 1.154.542 tweet dari tahun 2020 hingga 2023. Model deep learning BERT (IndoBERT) digunakan untuk mengklasifikasikan pandangan masyarakat Indonesia yang diperoleh dari media sosial X. Perbandingan model dengan parameter distribusi data training-validation-testing dilakukan untuk menentukan model terbaik. Model terbaik menggunakan distribusi data 70:20:10 dengan akurasi 93.60%, precision 93.64%, recall 93.60% dan f1-score 93.61% serta akurasi validasi sebesar 92.70%. Hasil analisis menunjukkan sentimen terhadap Korean Subculture terdiri dari 42.04% netral, 32.24% positif, dan 25.72% negatif. Drama Korea menjadi bagian Korean subculture yang paling diminati masyarakat Indonesia dibandingkan music dan makanan.

***Kata Kunci**: korean subculture, analisis sentimen, IndoBERT, drama korea, X (Twitter)*

Abstract

Korean subculture has experienced rapid growth in Indonesia in recent years. Indonesian society expresses positive, negative, or neutral views through social media X. (Twitter). Sentiment analysis is conducted to identify the public's views as positive, negative, or neutral. The stages of analysis include data collection, data pre-processing, data labeling, model training, classification, and visualization. The data analyzed consisted of 1,154,542 tweets from the years 2020 to 2023. The deep learning model BERT (IndoBERT) was used to classify the views of the Indonesian public obtained from social media X. A comparison of models with training-validation-testing data distribution parameters was conducted to determine the best model. The best model used a 70:20:10 data distribution with an accuracy of 93.60%, precision of 93.64%, recall of 93.60%, and an f1-score of 93.61%, as well as a validation accuracy of 92.70%. The analysis results show that sentiment towards Korean Subculture consists of 42.04% neutral, 32.24% positive, and 25.72% negative. Korean dramas are the most popular part of Korean subculture among the Indonesian public compared to music and food.

Keywords: korean subculture, sentiment analysis, IndoBERT, korean drama, X (Twitter)

1. Introduction

People now have easier access to a variety of Korean subculture content, such as K-pop, K-drama, and K-food, thanks to the digital age. Since the 1990s, the Korean subculture phenomenon has grown rapidly thanks to technology and the internet, particularly in Asia where cultures are close [1]. Teenagers in Indonesia enjoy the visual appeal, musicality, K-pop choreography, and the popularity of K-dramas through streaming platforms like Netflix and Viu [2]. Lifestyle trends such as collecting Korean-style clothing, mimicking language styles, and trying Korean cuisine influence many individuals. Public criticism of the Korean subculture phenomenon continues to arise even though many people find entertainment and lifestyle inspiration from it [3].

The popularity of Korean subculture has spurred business growth, particularly in the culinary sector, as public interest in trying Korean food has increased due to exposure from social media and Korean dramas [4]. Previous studies analyzed sentiment towards Korean dramas, K-pop idols, and Korean food using the Naïve Bayes algorithm or a combination with Support Vector Machine on social media data X. [5] [6] [4]. The sentiment analysis method is effective in understanding public perception of various social and cultural phenomena, including Korean subculture, with research results on seven dramas indicating a dominance of neutral sentiment. The drama Vincenzo received the highest

positive sentiment and is considered a favorite [5]. Research related to Korean food shows high public interest despite criticism regarding taste and halal standards, with a Naïve Bayes model accuracy of 70.97%. [4]. Analysis of K-pop idols such as BTS shows a majority of positive sentiment, with SVM having higher accuracy (81%) compared to Naïve Bayes (79%) [6].

Research using the BERT algorithm, such as IndoBERT, is becoming increasingly popular for sentiment analysis. The IndoBERT model achieved high accuracy of 98% on training data for Work From Home (WFH) sentiment during the COVID-19 pandemic [7]. Another study using IndoBERT to analyze application reviews shows high effectiveness, with an accuracy reaching 86–99% [8]. Analysis of hotel reviews using SmallBERT also revealed the majority of positive reviews (67.9%) with a model accuracy of 91.4% [9]. The new research uses data on Korean subculture (food, drama, music) from social media X and the IndoBERT algorithm to produce better analysis. The research stages include data collection, labeling, pre-processing, model training, testing, and visualization to provide in-depth insights into the Indonesian public's perception of Korean subculture.

2. Research Method

The research methodology uses an IndoBERT-based sentiment analysis approach to understand the perceptions of Indonesian society towards Korean subculture, including K-pop, K-drama, and Korean food. The data collection process was carried out through tweet harvesting on social media X (Twitter) and followed by stages of text pre-processing, labeling, model training, testing, and visualization. The IndoBERT-based approach aims to produce a better model and sentiment classification.

.

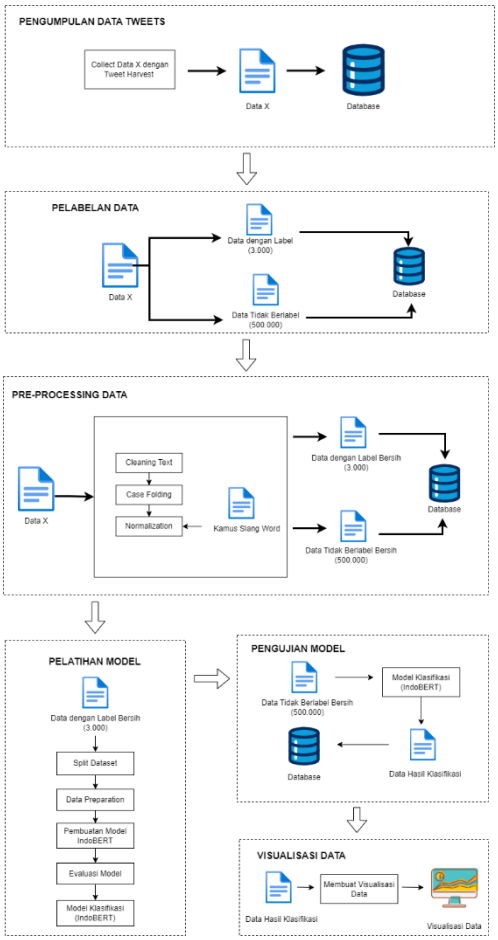

Figure 1. Overview of the Research

Figure 1 shows an overview of the research conducted, including several important steps. The data collection process was carried out through social media X, followed by the manual labeling of 3000 data points using three labels (positive, negative, and neutral), while the remaining data remained unlabeled for use in the testing phase. The pre-processing stage is carried out to clean the

tweet data, including text cleaning, case folding, normalization, and the removal of unclear words. The cleaned labeled data was trained using the pre-trained IndoBERT, while model testing was conducted to classify unlabeled data using the same algorithm. Visualizations were created based on the classification results obtained.

2.1 Pre-processing Data

The pre-processing process before sentiment analysis plays an important role in improving model accuracy and significantly impacts the analysis results [10]. The pre-processing stage aims to improve the quality of the data so that it is suitable for the classification process [11]. Pre-processing is carried out on all labeled and unlabeled data. The preprocessing stages include text cleaning, changing uppercase letters to lowercase, word normalization, and the removal of unclear words. The text cleaning process aims to remove unnecessary or unimportant parts from the data[12]. Cleaning text removes punctuation and other non-letter characters, as well as irrelevant symbols, hashtags, usernames, emoticons, and web addresses [10] [11].

Table 1. Text Cleaning Process

full_text	tweet_remove_char
@naraikha Gue punya playlist gitu teh, karena hobi banget dengerin lagu galau jadi kalo ditanyain selalu bingung favoritnya yang mana karena suka semua. Yang jelas lagu itu lagu kpop hehe. Kali ini mau rekomendasiin lagu ini aja! https://t.co/FXAoUStcSA	Gue punya playlist gitu teh karena hobi banget dengerin lagu galau jadi kalo ditanyain selalu bingung favoritnya yang mana karena suka semua Yang jelas lagu itu lagu kpop hehe Kali ini mau rekomendasiin lagu ini aja

Table 1 shows an example of the text cleaning process by displaying the text before and after it has been cleaned. The example shows that unnecessary characters, such as punctuation marks, mentions, and URLs like @naraikha and https://t.co/U1HmmS7dDA have been removed. The text cleaning process completed, the text goes through another pre-processing stage. The case folding process converts all uppercase letters to lowercase to standardize character casing. Case folding ensures that the data has a uniform letter format so that text analysis is not affected by the difference between uppercase and lowercase letters [12].

Table 2. Case Folding Process

tweet_remove_char	tweet_case_folding
Gue punya playlist gitu teh karena hobi banget dengerin lagu galau jadi kalo ditanyain selalu bingung favoritnya yang mana karena suka semua Yang jelas lagu itu lagu kpop hehe Kali ini mau rekomendasiin lagu ini aja	gue punya playlist gitu teh karena hobi banget dengerin lagu galau jadi kalo ditanyain selalu bingung favoritnya yang mana karena suka semua yang jelas lagu itu lagu kpop hehe kali ini mau rekomendasiin lagu ini aja

Table 2 shows an example of the process of converting uppercase letters to lowercase letters by displaying the text before and after the change. The example shows that the uppercase letters like "Gue" have been changed to lowercase "gue". The normalization process changes the spelling of a word using the slang dictionary that has been created. The slang dictionary comes from a source obtained through GitHub. The author added several words to the slang dictionary to adjust to the data used. The normalization process also removes unclear words like "wkwkwk" and "hahahaha" by matching them with the KBBI dictionary, standard words, the slang dictionary from GitHub, and the Korean dictionary. The Korean dictionary contains names of dramas, foods, groups, and other elements related to Korean subculture.

Table 3. Normalization and Removal of Ambiguos Words Process

tweet_case_folding	tweet_normalization	tweet_cleaned
gue punya playlist gitu teh karena hobi banget dengerin lagu galau jadi kalo ditanyain selalu bingung favoritnya yang mana karena suka semua yang jelas lagu itu lagu kpop hehe kali ini mau rekomendasiin lagu ini aja	saya punya playlist begitu teh karena hobi banget dengar lagu galau jadi kalau ditanyakan selalu bingung favoritnya yang mana karena suka semua yang jelas lagu itu lagu kpop hehe kali ini ingin menyarankan lagu ini saja	saya punya playlist begitu teh karena hobi banget dengar lagu galau jadi kalau ditanyakan selalu bingung favoritnya yang mana karena suka semua yang jelas lagu itu lagu kpop kali ini ingin menyarankan lagu ini saja

Table 3 shows an example of the normalization process that displays the text before and after being converted to standard spelling. The example shows that the word "gue" is changed to "saya," "dengerin" to "dengar," "kalo" to "kalau" and the unclear word "hehe" is removed. The normalization process is important to improve the quality of text data so that it can be analyzed better. The normalization process is important to improve the quality of text data so that it can be analyzed better.

2.2 BERT dan IndoBERT

Bidirectional Encoder Representations from Transformer (BERT) is a transformer-based machine learning approach developed by Google in 2018 by Jacob Devlin and his team. Google began using BERT in its search engine in 2019 and expanded its use to nearly all English-language queries by the end of 2020 [13]. The pre-trained model is trained through bidirectional learning by integrating context from both sides of the layer, allowing for fine-tuning by adding an additional layer for specific tasks [14]. Two paradigms of BERT training include pre-training and fine-tuning. The pre-training process involves unsupervised learning using large datasets such as BooksCorpus and English Wikipedia, while fine-tuning uses labeled data to adapt the model to specific tasks by adding task-specific layers [15]. IndoBERT is a BERT-based pre-training model designed for the Indonesian language, utilizing the Indo4B dataset which contains 4 billion words from various sources such as social media, blogs, news, and websites. IndoBERT uses SentencePiece with Byte Pair Encoding (BPE) as a tokenizer to form the vocabulary. IndoBERT is designed for general NLP tasks with limited computational resources [7].

3. Results and Analysis

The results and discussion elaborate on the training of the IndoBERT model with four data split scenarios, tokenization, and encoding processes to be accepted by IndoBERT, training results, evaluation of the best scenarios, and visualization of prediction results as well as trends in Korean subculture. The results of training the model with 4 scenarios produced accuracy, precision, recall, and f1-score values, and the model with the best results was used to predict sentiments that do not yet have labels. Data visualization displays a pie chart to show the overall prediction results and a line chart to observe trends related to Korean subculture.

3.1 Model Training and Evaluation

The division of labeled data before the training stage is done into three parts, namely training data, test data, and validation data. Training data is used for model development, test data is used to test the model and evaluate its accuracy, while validation data validates the model's performance and reduces the risk of overfitting. The sentiment analysis research on Korean subculture uses 3,000 labeled data for model development, while unlabeled data is utilized for sentiment classification.

The dataset used must conform to the input format accepted by IndoBERT. The tokenization process is carried out by adding special tokens such as [CLS] at the beginning of each sentence and [SEP] at the end of each sentence. The encoding process is carried out using a tokenizer based on the vocabulary index that has been previously trained on IndoBERT. The IndoBERT vocabulary is used in transfer learning, while words outside the vocabulary are split into subwords with the symbol ##. This input adjustment ensures that the data can be accepted by the IndoBERT model for further processing.

Table 4. IndoBERT Tokenization Process

Tweet Cleaned	sekarang ingin menulis tentang drama korea yang bagus menurut ku secret garden drama yang pertama kali banget aku lihat masih ingat sedih banget nonton ini terus jadi nagih menonton drama korea di sini kapten ri masih muda banget terus ha ji won dari dulu sampai sekarang masih cantik saja
Tokenization Results	['sekarang', 'ingin', 'menulis', 'tentang', 'drama', 'korea', 'yang', 'bagus', 'menurut', 'ku', 'secret', 'garden', 'drama', 'yang', 'pertama', 'kali', 'banget', 'aku', 'lihat', 'masih', 'ingat', 'sedih', 'banget', 'nonton', 'ini', 'terus', 'jadi', 'nag', '##ih', 'menonton', 'drama', 'korea', 'di', 'sini', 'kapten', 'ri', 'masih', 'muda', 'banget', 'terus', 'ha', 'ji', 'won', 'dari', 'dulu', 'sampai', 'sekarang', 'masih', 'cantik', 'saja']
Addition of Special Tokens	['[CLS]', 'sekarang', 'ingin', 'menulis', 'tentang', 'drama', 'korea',

	'yang', 'bagus', 'menurut', 'ku', 'secret', 'garden', 'drama', 'yang', 'pertama', 'kali', 'banget', 'aku', 'lihat', 'masih', 'ingat', 'sedih', 'banget', 'nonton', 'ini', 'terus', 'jadi', 'nag', '##ih', 'menonton', 'drama', 'korea', 'di', 'sini', 'kapten', 'ri', 'masih', 'muda', 'banget', 'terus', 'ha', 'ji', 'won', 'dari', 'dulu', 'sampai', 'sekarang', 'masih', '[SEP]']
Encoding	[2, 747, 561, 1971, 416, 4577, 3457, 34, 1305, 807, 425, 16369, 11710, 4577, 34, 736, 633, 2174, 304, 1173, 419, 3100, 5990, 2174, 5888, 92, 944, 472, 7805, 74, 5986, 4577, 3457, 26, 1585, 9919, 949, 419, 1734, 2174, 944, 2442, 2150, 7101, 98, 1248, 493, 747, 419, 3]
Attention Mask	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

The tokenization process utilizes the IndoBERT vocabulary without any additional modifications. Words in the example sentence are already present in the pre-trained model's vocabulary. The addition of special tokens such as [CLS] at the beginning and [SEP] at the end of the sentence is done before each token is encoded into the vocabulary index. Attention mask is used to distinguish between word tokens and padding, with a value of 1 for word tokens and 0 for padding.

The IndoBERT model was trained with the recommended hyperparameters, including a batch size of 16 or 32, a learning rate ranging from 5e-5 to 2e-5, and lasting for 2 to 4 epochs [14]. Fine-tuning the model is done by adding a classification layer using BertForSequenceClassification. The output token [CLS] is utilized to calculate logits and predictions through the softmax function. The model training uses the Adam optimizer with a learning rate of 0.00001, the SparseCategoricalCrossentropy loss function, and accuracy metrics. Training was conducted for two epochs with a batch size of 32. Validation data is used to monitor performance and prevent overfitting during training. The model predicts sentiment labels on the test data, and the evaluation results are presented in a classification report.

Table 5. IndoBERT Model Training Results

Scenario	Accuracy	Precision	Recall	F1-Score
80:10:10	92.33%	92.46%	92.33%	92.34%
70:20:10	**93.60%**	**93.64%**	**93.60%**	**93.60%**
70:15:15	92.67%	92.66%	92.67%	92.64%
60:20:20	92.83%	92.85%	92.83%	92.83%

Data splitting scenarios with ratios of 80:10:10, 70:15:15, 70:20:10, and 60:20:20 for training, validation, and testing. Changes in data distribution affect the model's performance based on the training results. The 80:10:10 ratio results in a model accuracy of 92.33% with consistent precision, recall, and f1-score. The 70:20:10 ratio increases the model accuracy to 93.60% due to a larger validation proportion. The 70:15:15 ratio results in a model accuracy of 92.67%, while the 60:20:20 ratio provides an accuracy of 92.83% and shows that the model remains reliable even with reduced training data.

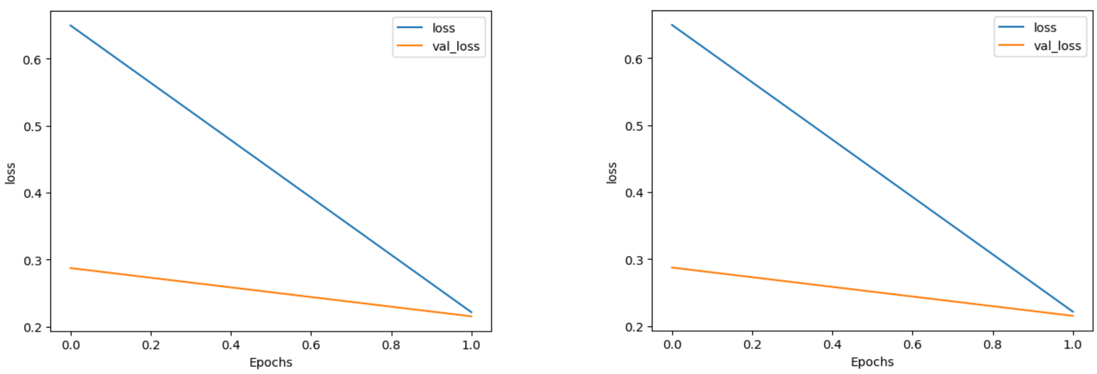

Figure 2. Accuracy and Loss Graph

Figure 2 displays the accuracy and loss graphs of the model with a 70:20:10 scenario. The accuracy graph on the left side shows an increase in model accuracy as the epochs progress. Training accuracy reaches over 92%, while validation accuracy approaches 91%. The increase in accuracy indicates the model's ability to learn effectively from the training data and maintain consistent performance on the validation data. The loss graph on the right side depicts a decrease in loss values for both training and validation data. The decrease in loss values indicates the model's improving ability to make correct predictions. The lower validation loss compared to the training loss reflects optimal performance in processing new data.

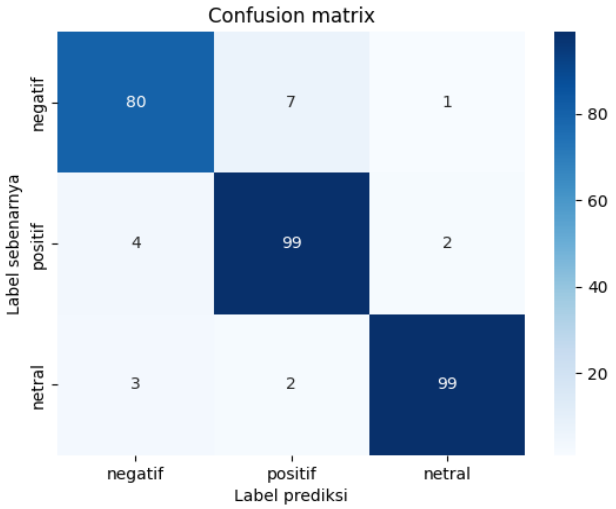

Figure 3. Confusion Matrix Diagram

Figure 3 shows the confusion matrix diagram for the 70:20:10 scenario. The first row shows the label “Negative”. The model correctly labeled 80 samples as “negative”, while 7 samples were labeled as “positive” and 1 sample as “neutral”. The second row shows the label “Positive”. A total of 99 samples have a “positive” label, with 4 samples labeled as “negative”, and 2 samples labeled as “neutral”. The third row shows the “Neutral” label. A total of 99 samples were correctly classified as “neutral”, while 3 samples as “negative” and 2 samples as “positive”.

3.2. Data Visualization

Data visualization simplifies large and complex data sets into easy-to-understand graphs. Visualization is used to analyze dataset composition, prediction results, growth trends, as well as the most frequently occurring words based on predictions. In addition, visualization helps identify hidden patterns in the data that can provide deeper insights.

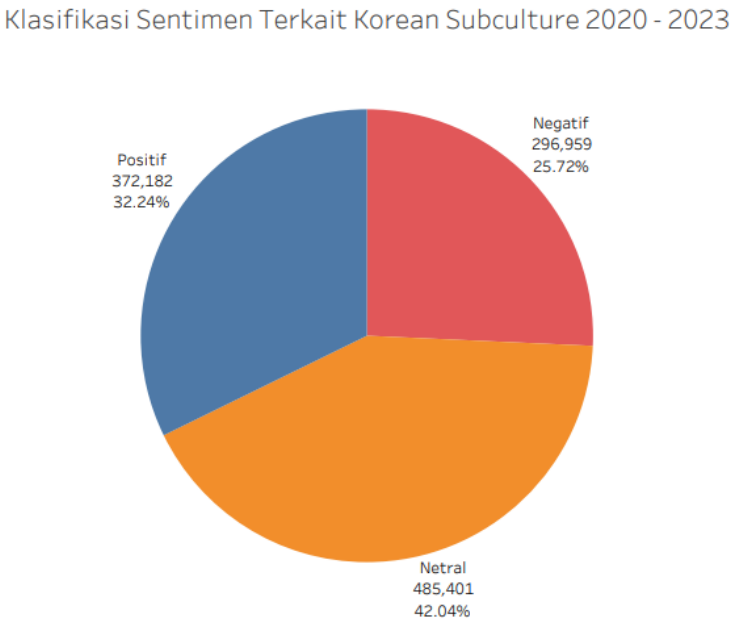

Figure 4. Pie Chart

Figure 4 shows the sentiment distribution of X users in Indonesia towards Korean Subculture (music, drama, and food) during 2020-2023. A total of 42.04% or 485,401 tweets have neutral sentiments, 32.24% or 372,182 tweets have positive sentiments, and 25.72% or 296,959 tweets have negative sentiments. The results show that Korean Subculture gets a lot of appreciation, there is also criticism. The majority of sentiment is neutral, indicating a response that tends to observe without giving a strong emotional reaction.

Figure 5. Line Chart

Figure 5 shows the trend of Korean Subculture during 2020-2021, covering Korean dramas, food, and music. The popularity of Korean dramas peaked in July and December 2020. The pandemic, along with the airing of It's Okay to Not Be Okay and Start Up, contributed to the increase in popularity on Netflix [16][17]. The downtrend occurred in November 2021 due to the absence of The Red Sleeve and Happiness from Netflix, the dominant platform in Indonesia, resulting in reduced audience access [18]. Korean music remains stable thanks to the regular activity of K-pop comebacks

every month. BTS, BLACKPINK, and NCT were active in 2020, while Aespa and Stray Kids appeared in 2021 [19]. Korean food trends are stagnant with no significant spikes, the stagnation reflects the steady level of acceptance of Korean food.

(a) Positive Sentiment Word Cloud

(b) Negative Sentiment Word Cloud

Figure 6. Word Cloud

Figure 6 displays the word cloud of positive and negative sentiments in November 2021 related to Korean dramas. Visualization (a) shows dominant words such as “cool”, “handsome”, “top”, and “addictive” indicating appreciation of the drama's quality, storyline, and acting. Terms such as “shock” and “effect” describe visual elements and surprises that attract the audience's attention. In contrast, visualization (b) shows negative sentiments with words such as “annoying”, “creepy”, “sin”, and “burden” reflecting criticism of the drama, e.g. heavy storyline or intense conflict. Terms such as “pity” and “heartbreak” highlight the tension or empathy the audience feels towards the story situation. In conclusion, although Korean dramas in November received a lot of appreciation, there were also criticisms related to aspects of the story, characters that were considered unsatisfactory and a variety of audience emotional responses to the genre of dramas that aired in November.

4. Conclusion

The study analyzed Indonesian sentiments towards Korean subculture covering music, drama, and food using 1,157,542 data from X social media. A total of 3,000 labeled data were used for IndoBERT model training, while the rest of the data were predicted and visualized. The best model with 70:20:10 data scenario achieved 93.60% accuracy, with precision, recall, and f1-score above 93%, and validation accuracy of 92.70%. Neutral sentiment dominated during the 2020-2023 period with a

proportion of 42.04%, followed by positive sentiment of 32.24% and negative sentiment of 25.72%. Korean dramas became the most popular category, especially with the presence of It's Okay to Not Be Okay and Start-Up. Korean music is showing stability thanks to the comeback activity of K-pop groups, while Korean food is stagnating, showing a consistent acceptance rate. Further research is recommended comparing other BERT models such as mBERT and XLM-R and applying various pre-processing methods to obtain a better model for sentiment analysis of Korean subculture. Expansion of the dataset by adding other aspects such as fashion, beauty, technology, and tourism is also recommended.

.

DAFTAR PUSTAKA

[1] A. H. Tama, "Analisis Fenomena Korean Wave Terhadap Sikap Fanatisme Pada Remaja Indonesia," *Jurnal Psimawa*, vol. 6, no. 1, pp. 1–5, 2023, [Online]. Available: http://jurnal.uts.ac.id/index.php/PSIMAWA

[2] D. Mawatdah, "Pengaruh Budaya K-PopTerhadap Perubahan Gaya Hidup Mahasiswa," Universitas Islam Negeri Ar-Raniry, Banda Aceh, 2022.

[3] R. Amelia, Darmansah, N. S. Prastiwi, and M. E. Purbaya, "Impementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 338, Apr. 2022, doi: 10.30865/jurikom.v9i2.3895.

[4] F. P. P. Subandi, F. Romadlon, I. Nurisusilawati, and A. Chindyana, "Sentiment Analysis of Indonesian Interest in Korean Food Based on Naïve Bayes Algorithm," *Jurnal Sosioteknologi*, vol. 21, no. 3, pp. 337–346, Dec. 2022, doi: 10.5614/sostek.itbj.2022.21.3.10.

[5] N. Raisa and N. Riza, "Sentimen Analisis Terhadap Opini Masyarakat Mengenai Drama Korea Pada Twitter Menggunakan Metode Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 2, pp. 1312–1320, 2023.

[6] R. Noviana and Rasal B A, "Penerapan Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Boy Band BTS pada Media Sosial Twitter," *JTS (Jurnal Teknik dan Science)*, vol. 2, no. 2, pp. 51–60, 2023.

[7] B. Rahmatullah, "Sentiment Analysis Pelaksanaan Work From Home Di Indonesia Pada Masa Pandemi Covid-19 Menggunakan IndoBERT," Institut Teknologi Sepuluh Nopember, Surabaya, 2021.

[8] K. P. W. R. M. R. Atmaja and W. Yustanti, "Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *JEISBI (Journal of Emerging Information Systems and Business Intelligence)*, vol. 02, no. 03, 2021.

[9] V. Chandradev, I. M. A. D. Suarjaya, and I. P. A. Bayupati, "Analisis Sentimen Review Hotel menggunakan Metode Deep Learning BERT," *Jurnal Buana Informatika*, vol. 14, no. 2, 2023.

[10] N. E. Oktaviana, Y. A. Sari, and Indriati, "Analisis Sentimen Terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features dan Support Vector Machine," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 9, no. 2, pp. 357–362, 2022, doi: 10.25126/jtiik.202295625.

[11] S. F. Pane and J. Ramdan, "Pemodelan Machine Learning : Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Menggunakan Data Twitter," *Jurnal Sistem Cerdas*, vol. 05, no. 01, pp. 12–20, 2022, [Online]. Available: https://t.co/IEnucGFuuJ

[12] M. F. Naufal and S. F. Kusuma, "Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis Deep Learning," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 1, pp. 44–49, Apr. 2022.

[13] D. N. L. P. V. Saraswati, N. Yudistira, and P. P. Adikara, "Analisis Sentimen terhadap Perundungan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 2, pp. 909–916, 2023, [Online]. Available: http://j-ptiik.ub.ac.id

[14] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *NAACL-HLT 2019*, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://github.com/tensorflow/tensor2tensor

[15] S. Rahmawati, “Implementasi Algoritma BERT Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Peduli Lindungi,” Universitas Teknologi Digital Indonesia, Yogyakarta, 2023.

[16] L. Mustinda, “It’s Okay to Not Be Okay Ada di Netflix, Ini Cara Streaming dan Downloadnya,” DetikInet. Accessed: Nov. 24, 2024. [Online]. Available: https://inet.detik.com/tips-dan-trik/d-5082028/its-okay-to-not-be-okay-ada-di-netflix-ini-cara-streaming-dan-downloadnya

[17] kumparanTECH, “Apa Itu Sandbox dalam Serial Start Up di Netflix?,” kumparanTECH. Accessed: Nov. 24, 2024. [Online]. Available: https://kumparan.com/kumparantech/apa-itu-sandbox-dalam-serial-start-up-di-netflix-1uhVrEIaKqT/full

[18] D. Angelia, “Platform Video Streaming Paling Digemari Masyarakat Indonesia 2022,” GoodStats. Accessed: Nov. 25, 2024. [Online]. Available: https://goodstats.id/article/platform-video-streaming-paling-digemari-masyarakat-indonesia-2022-qzfPB

[19] C. A. Ribeiro, “The 20 Best K-Pop Albums of 2021,” PopMatters. Accessed: Nov. 25, 2024. [Online]. Available: https://www.popmatters.com/best-k-pop-albums-2021