# Robust Sentiment Classification of Metaverse Services Using a Pre-trained Language Model with Soft Voting

**Haein Lee[1], Hae Sun Jung[2], Seon Hong Lee[1], and Jang Hyun Kim[3*]**
[1] Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, 03063, Korea
[e-mail: lhi00034, dltjsghd12@g.skku.edu]
[2] Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, 03063, Korea
[e-mail: jestiriel@g.skku.edu]
[3] Department of Interaction Science/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, 03063, Korea
[e-mail: alohakim@skku.edu]
[*]Corresponding author: Jang Hyun Kim

*Received May 1, 2023; revised August 6, 2023; accepted August 27, 2023; published September 30, 2023*

## Abstract

Metaverse services generate text data, data of ubiquitous computing, in real-time to analyze user emotions. Analysis of user emotions is an important task in metaverse services. This study aims to classify user sentiments using deep learning and pre-trained language models based on the transformer structure. Previous studies collected data from a single platform, whereas the current study incorporated the review data as "Metaverse" keyword from the YouTube and Google Play Store platforms for general utilization. As a result, the Bidirectional Encoder Representations from Transformers (BERT) and Robustly optimized BERT approach (RoBERTa) models using the soft voting mechanism achieved a highest accuracy of 88.57%. In addition, the area under the curve (AUC) score of the ensemble model comprising RoBERTa, BERT, and A Lite BERT (ALBERT) was 0.9458. The results demonstrate that the ensemble combined with the RoBERTa model exhibits good performance. Therefore, the RoBERTa model can be applied on platforms that provide metaverse services. The findings contribute to the advancement of natural language processing techniques in metaverse services, which are increasingly important in digital platforms and virtual environments. Overall, this study provides empirical evidence that sentiment analysis using deep learning and pre-trained language models is a promising approach to improving user experiences in metaverse services.

***Keywords:*** BERT, metaverse, natural language processing, pre-trained language model, ubiquitous computing

**http://doi.org/10.3837/tiis.2023.09.002** **ISSN : 1976-7277**

## 1. Introduction

Ubiquitous computing is frequently utilized to document events from a social, physical, and economic point of view. In particular, data of ubiquitous computing mainly consists of the information generated by geographical factors and time stamps of samples sequentially observed from mobile and human sensing devices or simulations in various fields [1]. With advancements in big data, many researchers and industries have focused on the technology required to process it. Web pages and reviews are examples of unstructured data that comprise most digital information. Millions of such text data are continually being generated as streaming data. Hence, natural language processing (NLP) tasks such as event detection in documents or calculating similarities must be promptly processed to obtain correct information [2]. Therefore, tasks related to NLP face considerable challenges in terms of scalability in text-data processing.

One area where ubiquitous computing has had a significant impact is in mobile application usage, which has become an integral part of daily life for many people worldwide [3]. People worldwide spend times watching videos via different video applications such as YouTube, Twitch, and TikTok [4]. These platforms allow them to share their opinions about a video through the comments function. Several researchers have attempted to conduct sentiment analysis based on the data gathered with these comments. However, in previous studies, despite the large amount of comment data generated by YouTube, only a limited amount could be utilized using the YouTube API. Additionally, because the analysis was performed using data from a single platform, there was difficulty in examining large amounts of data from users across multiple platforms [5–7].

Another common activity among mobile users is socialization through applications. As the metaverse expands its horizons, different technologies are being adopted by an increasing number of platforms and applications in their services [8]. Moreover, many users experience diverse services such as learning, travel, and lectures through metaverse applications. However, few studies have been undertaken on mobile applications that support metaverse services. As the number of users and companies that use metaverse services in mobile applications increases, it is crucial to analyze user experiences based on big data collected from users to provide improved services [9].

This study presents an improved classification model that can analyze both the user reviews and comments of mobile applications generated in real-time from YouTube videos. The domain for the metaverse service was specified, and the number of data points was 450,890. YouTube video data were collected using "Metaverse" as a keyword. Additionally, four mobile applications that provide metaverse services were selected from the Google Play Store: Roblox, Minecraft, Animal Crossing, and Zepeto. These application services have recorded a maximum of 150 thousand monthly users since they began and are highly successful [10–12]. The authors implemented the Bidirectional Encoder Representations from Transformers (BERT) model, A Lite BERT (ALBERT), and Robustly optimized BERT approach (RoBERTa) model with an ensemble model, enabling them to build a sentiment classification model. This study is noteworthy because it combines natural language data from diverse platforms in the metaverse domain and suggests a novel deep learning approach.

# 2. Related Works

## 2.1 Applications of Machine Learning in Various Domains

With the development of machine-learning technology, research on classification and prediction has been conducted in various domains. **Table 1** shows a summary of related research conducted in diverse domains. Lee et al. [13] used algorithms to predict customer satisfaction for mobile health services and demonstrated that extreme gradient boosting (XGBoost) can be adopted to forecast users' sentiments. To predict sentiment using restaurant feedback on quality and price, Zahoor et al. [14] confirmed that Random Forest algorithm has higher accuracy than other algorithms. Elnagar et al. [15] achieved improved performance in the categorization of Arabic text using deep learning methodologies (i.e., convolutional-Gated Recurrent Unit (GRU), Attention-GRU). Additionally, Jung et al. [16] predicted Bitcoin price trends using data from Reddit and LexisNexis over a period of more than four years. The study found that the use of extreme gradient boosting (XGBoost) and a combination of sentiment and technical indicators can produce significant results with an accuracy of 90.57% and an area under the receiver operating characteristic curve value (AUC) of 97.48%. Munisami et al. [17] presented a recognition system that identifies plants based on images of their leaves, which can be useful for various professionals. The system utilized pre-processing, feature extraction techniques, and a pattern matcher, achieving an accuracy of 83.5% which improved to 87.3% with the use of color histogram information. Yang et al. [18] designed a fault location method based on deep neural evolution network (FL-DNEN) to extract hidden fault characteristics from alarm information collected from cloud data center, in order to address the vulnerabilities that cloud computing faces due to network failures. By adopting this approach, they achieved a fault location accuracy of up to 92%. Lee et al. [19] leveraged extensive data from news articles and academic papers to analyze Environmental, Social and Governance (ESG) discourse. Employing the Bidirectional Encoder Representations from Transformers (BERTopic), the researchers identified significant keywords, and tracked the evolution of topic patters over time. Yang et al. [20] conducted a comprehensive assessment of various core-level influencing factors to evaluate the quality of link transmission. Additionally, the authors introduced an unsupervised learning-based modeling methodology, a self-organizing feature mapping (SOFM)-based routing, and spectrum allocation algorithm. Furthermore, the application of machine learning techniques is being widely utilized in metaverse services and environments. Moztarzadeh et al. [21] confirmed the applicability of disease digitalization using machine learning within the metaverse environment. Specifically, the gradient boosting algorithm was recognized as a potent and efficient approach for digitizing cancer. Tunca et al. [22] aimed to facilitate the understanding of the metaverse. The authors sought to grasp the positive and negative features of emotions. In addition, the correlations with the metaverse concept and related concepts was elucidated. For this purpose, they employed unsupervised machine learning-based Leximancer and Valence Aware Dictionary and sEntiment Reasoner (VADER) for natural language data analysis. Abou El-Magd et al. [23] designed a classification model using machine learning and real-world data to enhance the interaction between the metaverse and physical reality. They achieved a performance of 99.9% using SVM's three-dimension kernel model. The performance varies depending on the domain. Therefore, many studies are being conducted to confirm the model's capability in each field.

**Table 1.** Summary of Machine Learning Research in Diverse Domains

| **Authors** | **Data** | **Model** | **Result (Accuracy)** |
|---|---|---|---|
| Lee et al. (2022) [13] | Mobile healthcare application (Samsung Health) reviews | Logistic Regression<br>Random Forest<br>XGBoost<br>Gradient Boost<br>Naïve Bayes | 0.891<br>0.8662<br>0.8916<br>0.8896<br>0.8234 |
| Zahoor et al. (2020) [14] | Restaurant reviews of Karachi in Pakistan | Naïve Bayes<br>Logistic Regression<br>Support Vector Machine<br>Random Forest | 84%<br>86%<br>89%<br>94% |
| Elnagar et al. (2020) [15] | Arabic news articles datasets (Single-label and Multi-label) | convolutional-GRU<br>attention-GRU<br>attention-GRU | 91.18%; Single-label<br>96.94%; Single-label<br>88.68%; Multi-label |
| Jung et al. (2023) [16] | Reddit and LexisNexis data | Logistic Regression<br>Naïve Bayes<br>Support Vector Machine<br>Random Forest<br>XGBoost<br>LightGBM | 0.808510<br>0.580547<br>0.784194<br>0.854103<br>0.905775<br>0.893617 |
| Munisami et al. (2015) [17] | Folio (Images of leaves captured from the University of Mauritius farm) | K-Nearest Neighbor | 87.3% |
| Yang et al. (2020) [18] | Alarm set and scope of faults | FL-DNEN<br>FL-DNN<br>FL-SVM | 92%<br>81%<br>75% |
| Abou El-Magd et al. (2023) [23] | Beef quality monitoring dataset using an e-nose | Linear SVM<br>Quadratic SVM<br>Cubic SVM | 97.3%<br>99.7%<br>99.9% |

## 2.2 Sentiment Analysis Utilizing Social Network Services Data

Sentiment analysis, also referred to opinion mining, analyzes a person's thoughts and feelings using natural language processing techniques [24]. Recently, many researchers have demonstrated an understanding in classifying emotions. In particular, different social network services, including blogs, reviews, and articles, have been employed to obtain the general public's thoughts on a specific topic or circumstance. Lee et al. [25] indicated that an algorithm combining VADER and machine learning is an effective methodology for sentiment analysis using metaverse mobile application service reviews. La et al. [26] found that logistic regression is an appropriate model for analyzing the online comments of tourists. Luo and Xu

[27] used online restaurant reviews to discover the model that could predict positive or negative scores. Novendri et al. [28] adjusted YouTube comments to build sentiment classifiers and achieved successful results using Naïve Bayes. Hence, collecting and leveraging user reviews is the most advantageous approach for researchers to understand user opinions and feelings.

## 3. Method

In this study, the authors considered a robust deep learning model that could be used on various platforms. Previous studies normally trained a deep learning model using the data collected from a single platform to classify sentiments. However, this approach has its limitations with regard to generalization. To alleviate this shortcoming, review data from YouTube and Google Play Stores were used to train the deep learning model. The flow of architecture of the proposed system is illustrated in **Fig. 1**.

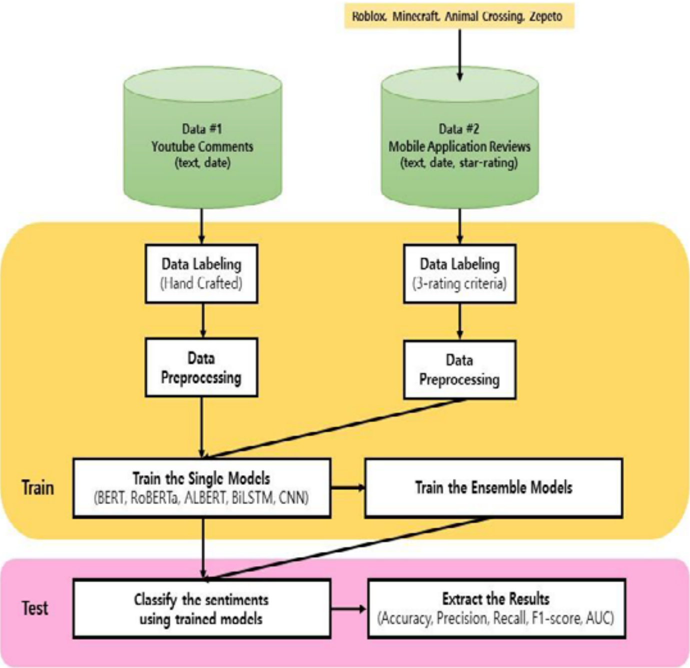


**Fig. 1.** Architecture of the total process.

### 3.1 Data Collection

The authors collected 49,306 review data from 18 videos posted on YouTube using "metaverse" as a keyword. More specifically, YouTube videos were selected from platforms publishing news, documentaries, and technical information, as well as from individual YouTubers who introduced technology and issues. Moreover, 401,584 review data were collected from four applications supporting metaverse services: Roblox, Minecraft, Animal Crossing, and Zepeto in the Google Play Store. the number of data points was 450,890 (**Table 2**) and examples of the data used can be found in **Table 3**.

**Table 2.** Information on the collected data

| Data Source | Topic | The number of data | Date |
|---|---|---|---|
| Youtube | Metaverse | 49,306 | Mar. 01. 2021 - Mar. 02. 2022 |
| Roblox (Google Play Store) | Metaverse | 85,332 | Jun. 25. 2014 - Nov. 23. 2021 |
| Minecraft (Google play store) | Metaverse | 81,695 | Mar. 05. 2021 - Nov. 16. 2021 |
| Animal Crossing (Google play store) | Metaverse | 117,375 | Oct. 25. 2017 - Nov. 16. 2021 |
| Zepeto (Google play store) | Metaverse | 117,182 | Sep. 19. 2018 - Nov. 16. 2021 |

**Table 3.** Examples of collected data

<table>
<tr><th colspan="4">Youtube</th></tr>
<tr><th colspan="2">Text</th><th>Label</th><th>Date</th></tr>
<tr><td colspan="2">The metaverse can be so much more. The only constraints are our own imaginations</td><td>1 (positive)</td><td>Feb. 2022</td></tr>
<tr><td colspan="2">This is by far the most inspiring thing ive ever seen…im inspired to throw away every electronic I own and run to live in the forest</td><td>2 (negative)</td><td>Nov. 2021</td></tr>
<tr><td colspan="2">whats the difference between a VR that looks and feels exactly like the real world and the real world?</td><td>3 (neutral)</td><td>Feb. 2022</td></tr>
<tr><th colspan="4">Google Play Store</th></tr>
<tr><th>Category</th><th>Text</th><th>Rating (1 to 5)</th><th>Date</th></tr>
<tr><td>Roblox</td><td>Doesn’t work</td><td>1</td><td>May. 2019</td></tr>
<tr><td>Minecraft</td><td>The worst game ever</td><td>1</td><td>Nov. 2021</td></tr>
<tr><td>Animal Crossing</td><td>Really good and super cute</td><td>5</td><td>Nov. 2021</td></tr>
<tr><td>Zepeto</td><td>Nice game</td><td>5</td><td>Nov. 2021</td></tr>
</table>

### 3.2 Data Preprocessing

The authors labeled the data to classify the positive, negative, and neutral classes in the YouTube review data. Subsequently, the data from the neutral class were excluded from this study. Existing studies indicate that reviews with four- or five-star ratings out of five were considered positive, whereas those with less than three stars were considered negative [29]. Additionally, the authors removed unnecessary duplication of data.

Most datasets have differences in the number of data points that belong to each class. This disproportion can degrade the model performance. To reduce the dissymmetry, down sampling was utilized to balance the number of classes [30,31]. To train a deep learning model, a training dataset and a test dataset were randomly separated from the entire dataset in a ratio of 8:2. Subsequently, to avoid overfitting, 20% of the training dataset was used as the validation dataset.

### 3.3 Deep Learning Model and Pre-trained Language Model

Various deep learning models have been used to solve NLP tasks. Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are representative deep learning models. In this study, a CNN and Bidirectional LSTM (BiLSTM) were used. In addition, after the advent of the transformer model, there have been many cases in which natural language processing tasks have been solved by fine-tuning them with pre-trained language models using the transformer. The authors applied representative pre-trained models, BERT, RoBERTa, and ALBERT, in HuggingFace with appropriate tokenizers.

#### 3.3.1 Bidirectional Encoder Representations from Transformers (BERT)

The BERT model was created using the encoder part of the transformer. With large-capacity text data, a pre-trained bidirectional language model (LM) through various tasks, such as masked LM and next sentence prediction (NSP) [32]. As a result, BERT shows strength in natural language understanding tasks. Our study utilized classification among tasks that could be performed using BERT.

#### 3.3.2 Robustly Optimized BERT Approach (RoBERTa)

RoBERTa is an enhanced model of BERT that generally performs better than it. The training dataset used in BERT was 16GB of data from BookCorpus and WIKIPEDIA. However, RoBERTa employed 160GB of the training dataset in five domains with different capacities. In addition, RoBERTa does not apply NSP, which is one of the BERT operations used for training. While BERT used a static mask pattern for every epoch, RoBERTa implemented the various mask patterns using the dynamic mask technique [33].

#### 3.3.3 A Lite BERT (ALBERT)

ALBERT, the lite version of BERT, reduces the number of parameters while minimizing the performance degradation of BERT using two methods. First, a linear layer that reduces the dimension of the embedding layer through factorized embedding parameterization and restores it to an existing scale of a hidden size is added. Second, the parameters of the attention and feed-forward network (FFN) layers were shared through cross-layer parameter sharing. Additionally, because the NSP was not helpful in improving the performance, Sentence Order Prediction (SOP), a more difficult task, was adopted [34].

### 3.4 Ensemble Classifier

After training the single models, the authors configured the voting classifiers. There are two types of voting classifiers: hard voting classifiers that predict using a majority vote and soft voting classifiers that average the predicted probabilities obtained from each model and select the best class [35,36]. This study constructed ensemble models employing soft voting. Because the soft voting classifier combines predictions from different models, it usually achieves better outcomes than the other base models [37]. Therefore, the proposed study designed several ensemble models for all possible combinations using deep-learning models: BERT, RoBERTa, and ALBERT.

## 4. Performance Evaluation

The experimental setup and classification assessment are presented in this section. The authors experimentally verified the use of ensemble deep-learning models.

### 4.1 Evaluation Methodology

Ten deep learning classifiers were used for the verification. The evaluation was conducted by adjusting the confusion matrix that dealt with the accuracy (1), recall (2), precision (3), and F1-score (4) [38]. Additionally, the authors measured the area under the curve (AUC), which correlates to the section under the receiver operating characteristic (ROC) curve. AUC is a metric indicating how well a classifier distinguishes classes; the closer the score is to 1, the better the performance [39]. **Table 4** defines the confusion matrix.

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN) \quad (1)$$

$$\text{Recall} = TP / (TP+FN) \quad (2)$$

$$\text{Precision} = TP / (TP+FP) \quad (3)$$

$$\text{F1-score} = 2 \times (\text{Recall} \times \text{Precision}) / (\text{Recall}+\text{Precision}) \quad (4)$$

**Table 4.** Confusion matrix

| Confusion matrix | | Real | |
|---|---|---|---|
| | | Positive | Negative |
| Prediction | Positive | True Positive (TP) | False Positive (FP) |
| | Negative | False Negative (FN) | True Negative (TN) |

### 4.2 Results

The results show two conspicuous ensemble models, as depicted in **Table 5** and **Figs. 2-4**. As can be seen, the first (BERT and RoBERTa) and second ensemble models (BERT, RoBERTa, and ALBERT) both demonstrate high accuracies (88.57% and 88.47%, respectively). In addition, the AUC score of the ensemble model combining the above three models was 0.9458, which is close to 1. Therefore, the ensemble model employing RoBERTa performed better than the other combinations.

**Table 5.** Results of the deep learning models

| Model | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| BERT | 0.877920 | 0.858241 | 0.905387 | 0.881183 | 0.941533 |
| ALBERT | 0.873648 | 0.852177 | 0.904132 | 0.877386 | 0.938126 |
| RoBERTa | 0.882988 | 0.865655 | 0.906690 | 0.885698 | 0.942812 |
| BERT + ALBERT | 0.879947 | 0.860261 | 0.907269 | 0.883140 | 0.943026 |
| BERT + RoBERTa | 0.885740 | 0.868826 | 0.908669 | 0.888301 | 0.945733 |

| | | | | | |
|---|---|---|---|---|---|
| ALBERT + RoBERTa | 0.884002 | 0.866151 | 0.908379 | 0.886763 | 0.945105 |
| BERT + ALBERT + RoBERTa | 0.884702 | 0.865959 | 0.910310 | 0.887581 | 0.945889 |
| BiLSTM | 0.834113 | 0.832109 | 0.837130 | 0.834612 | 0.913482 |
| CNN | 0.839906 | 0.812081 | 0.884485 | 0.846738 | 0.911113 |
| BILSTM + CNN | 0.851008 | 0.825157 | 0.890760 | 0.856705 | 0.920718 |

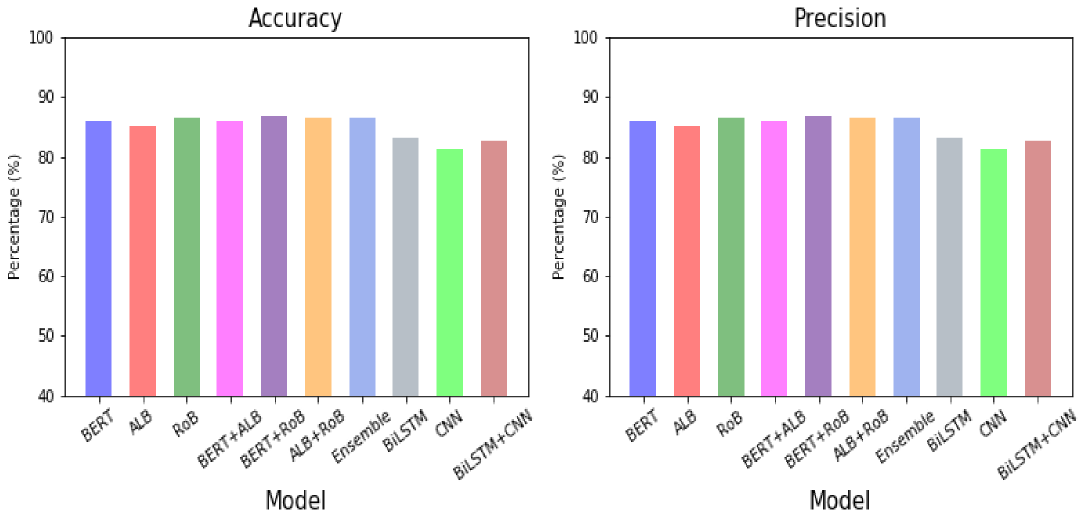


**Fig. 2.** Results of accuracy and precision
**Note.** (ALB: ALBERT, RoB: RoBERTa, Ensemble: BERT + ALBERT + RoBERTa).

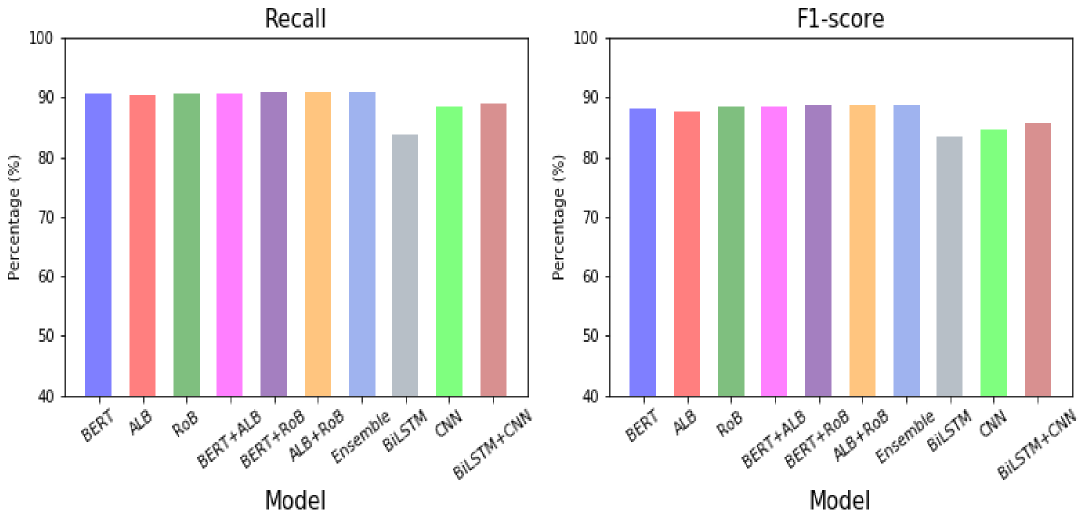


**Fig. 3.** Results of recall and F1-score
**Note.** (ALB: ALBERT, RoB: RoBERTa, Ensemble: BERT + ALBERT + RoBERTa).

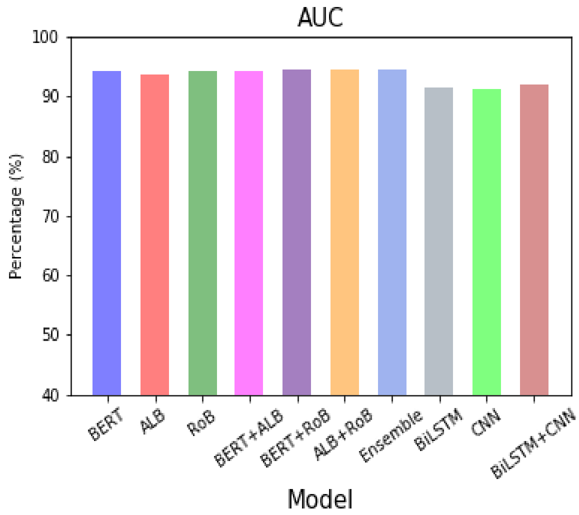


**Fig. 4.** Results of AUC
**Note.** (ALB: ALBERT, RoB: RoBERTa, Ensemble: BERT + ALBERT + RoBERTa).

## 5. Conclusion

The growth of mobile technology and the internet has led to the creation of big data in real-time across many industries. Many researchers have focused on technology that processes big data [1]. The majority of big data consists of text data. Therefore, text mining in data analytics is emerging as a powerful technique for extracting new information and classifying latent meanings from unstructured text data [40]. In addition, the increasing availability of online text data has created opportunities for an advanced understanding of customer preferences based on user-generated content [41]. In this context, sentiment analysis research that examines user experience has significantly improved user satisfaction.

In previous studies that attempted sentiment analysis using YouTube comment data, the analysis was implemented with a small amount of data [5–7]. Furthermore, when using review data from mobile applications, single-platform data were employed [9,25]. One limitation is that this model was not robust because of the lack of diversity. In this study, the authors provide deep learning models that can process big data collected from multiple platforms (i.e., YouTube and four mobile applications) for a specific domain. In addition, the sentiment classifier was evaluated by applying ten different models. As a result, the ensemble of BERT and RoBERTa showed a high accuracy of 88.57%.

Following the practical implications, two types of ensemble models performed well: the first ensemble model (BERT and RoBERTa) and the second ensemble model (BERT, RoBERTa, and ALBERT). These findings could be applied generally in domains relevant to metaverse services, since they were obtained using large amounts of data taken from many platforms. Additionally, in industries with limited computing resources, it is recommended to perform user analysis based on the RoBERTa model because the overall performance is good when combined with the RoBERTa model. In summary, the authors developed a classifier that performed well in the metaverse domain. Thus, the results of this study are useful for investigating how consumers perceive metaverse-related services.

From a theoretical perspective, the authors confirmed that models utilizing user-generated content such as online reviews can predict user satisfaction. Online reviews are natural language data generated by various sectors, suggesting that they can be used to improve services. Therefore, this study reinforces a novel methodological addition to existing literature on metaverse services.

The following are this study's limitations: First, only English data were used. Therefore, further studies that consider other languages and cultures should be undertaken. Second, because the data were collected using metaverse keywords, our discovery can only be used for metaverse services. Therefore, other fields utilizing them require additional validation.

## Acknowledgements

This study was supported by a National Research Foundation of Korea (NRF) (http://nrf.re.kr/eng/index) grant funded by the Korean government (RS-2023-00208278).
We appreciate Editage (www.editage.co.kr) for their English editing service.

## References

[1] C. Yang, K. Clarke, S. Shekhar and C. V. Tao, "Big Spatiotemporal Data Analytics: A research and innovation frontier," *International Journal of Geographical Information Science*, vol. 34, no. 6, pp. 1075-1088, 2020. Article (CrossRef Link)

[2] R. Agerri, X. Artola, Z. Beloki, G. Rigau and A. Soroa, "Big data for Natural Language Processing: A streaming approach," *Knowledge-Based Systems*, vol. 79, pp. 36-42, 2015. Article (CrossRef Link)

[3] A. A. L. Cunha, M. C. Costa and M. A. C. Pacheco, "Sentiment analysis of YouTube video comments using deep neural networks," in *Proc. of International Conference on Artificial Intelligence and Soft Computing (ICAISC)*, Zakopane, Poland, pp. 561-570, 2019. Article (CrossRef Link)

[4] Statista, "Most popular mobile video apps worldwide," 2021. [Online]. Available: https://www.statista.com/statistics/1251175/top-video-apps-worldwide/

[5] A. U. R. Khan, M. Khan and M. B. Khan, "Naïve Multi-label classification of YouTube comments using comparative opinion mining," *Procedia Computer Science*, vol. 82, pp. 57-64, 2016. Article (CrossRef Link)

[6] R. Benkhelifa and F. Z. Laallam, "Opinion extraction and classification of real-time youtube cooking recipes comments," in *Proc. of International Conference on Advanced Machine Learning Technologies and Applications (AMLTA)*, Cairo, Egypt, pp. 395-404, 2018. Article (CrossRef Link)

[7] M. Aufar, R. Andreswari and D. Pramesti, "Sentiment analysis on youtube social media using decision tree and random forest algorithm: A case study," in *Proc. of 2020 International Conference on Data Science and Its Applications (ICoDSA)*, Bandung, Indonesia, pp. 1-7, 2020. Article (CrossRef Link)

[8] S. N. Suzuki, H. Kanematsu, D. M. Barry, N. Ogawa, K. Yajima et al., "Virtual Experiments in Metaverse and their Applications to Collaborative Projects: The framework and its significance," *Procedia Computer Science*, vol. 176, pp. 2125-2132, 2020. Article (CrossRef Link)

[9] M. Yavuz, E. Çorbacıoğlu, A. N. Başoğlu, T. U. Daim and A. Shaygan, "Augmented reality technology adoption: Case of a mobile application in Turkey," *Technology in Society*, vol. 66, pp. 101598, 2021. Article (CrossRef Link)

[10] P. A. Rospigliosi, "Metaverse or Simulacra? Roblox, Minecraft, Meta and the turn to virtual reality for education, socialisation and work," *Interactive Learning Environments*, vol. 30, no. 1, pp. 1-3, 2022. Article (CrossRef Link)

[11] J. C. Fisher, N. Yoh, T. Kubo and D. Rundle, "Could Nintendo's Animal Crossing be a tool for conservation messaging?," *People and Nature*, vol. 3, no. 6, pp. 1218-1228, 2021. Article (CrossRef Link)
[12] BUSINESS 2 COMMUNITY, "Best Metaverse Apps for 2022," 2022. [Online]. Available: https://www.business2community.com/cryptocurrency/best-metaverse-apps
[13] H. Lee, S. H. Lee, D. Nan and J. H. Kim, "Predicting User Satisfaction of Mobile Healthcare Services Using Machine Learning: Confronting the COVID-19 Pandemic," *Journal of Organizational and End User Computing (JOEUC)*, vol. 34, no. 6, pp. 1-17, 2022. Article (CrossRef Link)
[14] K. Zahoor, N. Z. Bawany and S. Hamid, "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning," in *Proc. of 2020 21st International Arab Conference on Information Technology (ACIT)*, Ajman, UAE, pp. 1-6, 2020. Article (CrossRef Link)
[15] A. Elnagar, R. Al-Debsi and O. Einea, "Arabic text classification using deep learning models," *Information Processing & Management*, vol. 57, no. 1, pp. 102121, 2020. Article (CrossRef Link)
[16] H. S. Jung, S. H. Lee, H. Lee and J. H. Kim, "Predicting bitcoin trends through machine learning using sentiment analysis with technical indicators," *Computer Systems Science and Engineering*, vol. 46, no.2, pp. 2231–2246, 2023. Article (CrossRef Link)
[17] T. Munisami, M. Ramsurn, S. Kishnah and S. Pudaruth, "Plant leaf recognition using shape features and colour histogram with K-nearest neighbour classifiers," *Procedia Computer Science*, vol. 58, pp. 740–747, 2015. Article (CrossRef Link)
[18] H. Yang, X. Zhao, Q. Yao, A. Yu, J. Zhang and Y. Ji, "Accurate fault location using deep neural evolution network in cloud data center interconnection," *IEEE Transactions on Cloud Computing*, vol. 10, no. 2, pp. 1402-1412, 2022. Article (CrossRef Link)
[19] H. Lee, S. H. Lee, K. R. Lee and J. H. Kim, "ESG Discourse Analysis Through BERTopic: Comparing News Articles and Academic Papers," *Computers, Materials & Continua*, vol. 75, no. 3, pp. 6023-6037, 2023. Article (CrossRef Link)
[20] H. Yang, Q. Yao, B. Bao, A. Yu, J. Zhang and A. V. Vasilakos, "Multi-associated parameters aggregation-based routing and resources allocation in multi-core elastic optical networks," *IEEE/ACM Transactions on Networking*, vol. 30, no. 5, pp. 2145-2157, 2022. Article (CrossRef Link)
[21] O. Moztarzadeh, M. Jamshidi, S. Sargolzaei, A. Jamshidi, N. Baghalipour, M. Malekzadeh Moghani and L. Hauer, "Metaverse and Healthcare: Machine Learning-Enabled Digital Twins of Cancer," *Bioengineering*, vol. 10, no. 4, pp. 455, 2023. Article (CrossRef Link)
[22] S. Tunca, B. Sezen and V. Wilk, "An exploratory content and sentiment analysis of the guardian metaverse articles using leximancer and natural language processing," *Journal of Big Data*, vol. 10, no. 1, pp. 82, 2023. Article (CrossRef Link)
[23] [L. M. Abou El-Magd, N. M. Abdelnapi, A. Darwish and A. E. Hassanien, "A Proposed Metaverse Framework for Food Security Based-IoT Network and Machine Learning," in *The Future of Metaverse in the Virtual Era and Physical World*, vol. 123, Cha, Springer International Publishing, 2023, pp. 137-153.
[24] M. Singh, A. K. Jakhar and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Social Network Analysis and Mining*, vol. 11, no. 1, pp. 1-11, 2021. Article (CrossRef Link)
[25] S. H. Lee, H. Lee and J. Hyun Kim, "Enhancing the prediction of user satisfaction with metaverse service through machine learning," *Computers, Materials & Continua*, vol. 72, no.3, pp. 4983–4997, 2022. Article (CrossRef Link)
[26] L. La, F. Xu, Q. Ren, F. Zhen and T. Lobsang, "Understanding consumers' sentiment expressions in online reviews: a hybrid approach," *Journal of International Consumer Marketing*, vol. 34, no. 4, pp. 465-479, 2022. Article (CrossRef Link)
[27] Y. Luo and X. Xu, "Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic," *International Journal of Hospitality Management*, vol. 94, pp. 102849, 2021. Article (CrossRef Link)

[28] R. Novendri, A. S.Callista, D. N. Pratama and C. E. Puspita, "Sentiment analysis of YouTube movie trailer comments using naïve bayes," *Bulletin of Computer Science and Electrical Engineering*, vol. 1, no. 1, pp. 26-32, 2020. Article (CrossRef Link)

[29] K. K. Mohbey, "Sentiment analysis for product rating using a deep learning approach," in *Proc. of 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Tamil Nadu, India, pp. 121-126, 2021. Article (CrossRef Link)

[30] M. A. Bashar and R. Nayak, "Active learning for effectively fine-tuning transfer learning to downstream task," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 12, no. 2, pp. 1-24, 2021. Article (CrossRef Link)

[31] K. Khandagale and H. Gandhi, "Sarcasm Detection in Hindi-English Code-Mixed Tweets Using Machine Learning Algorithms," in *Proc. of International Conference on Computing in Engineering & Technology*, Singapore, pp. 221-229, 2022. Article (CrossRef Link)

[32] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018. Article (CrossRef Link)

[33] Y. Liu, M. Ott, N. Goyal, J. Du, M Joshi et al., "Roberta: A robustly optimized bert pretraining approach," 2019. Article (CrossRef Link)

[34] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma et al., "Albert: A lite bert for self-supervised learning of language representations," 2019. Article (CrossRef Link)

[35] S. Mishra, P. K. Mallick, H. K. Tripathy, L. Jena and G. S. Chae, "Stacked KNN with hard voting predictive approach to assist hiring process in IT organizations," *The International Journal of Electrical Engineering & Education*, p. 0020720921989015, 2021. Article (CrossRef Link)

[36] H. Wang, Y. Yang, H. Wang and D. Chen, "Soft-voting clustering ensemble," in *Proc. In International Workshop on Multiple Classifier Systems*, Nanjing, China, pp. 307-318, 2013. Article (CrossRef Link)

[37] S. Kumari, D. Kumar and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40-46, 2021. Article (CrossRef Link)

[38] J. T. Townsend, "Theoretical analysis of an alphabetic confusion matrix," *Perception & Psychophysics*, vol. 9, no. 1, pp. 40-50, 1971. Article (CrossRef Link)

[39] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Transactions on knowledge and Data Engineering*, vol. 17, no. 3, pp. 299-310, 2005. Article (CrossRef Link)

[40] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani and M. R. Yeganegi, "Text mining in big data analytics," *Big Data and Cognitive Computing*, vol. 4, no. 1, pp. 1, 2020. Article (CrossRef Link)

[41] D. Mittal and S. R. Agrawal, "Determining banking service attributes from online reviews: text mining and sentiment analysis," *International Journal of Bank Marketing*, vol. 40, no. 3, pp. 558-577, 2022. Article (CrossRef Link)

**Haein Lee** is PhD student at Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University. Her research focuses on Natural Language Processing, Deep Learning, and Machine Learning.

**Hae Sun Jung** is PhD student at Department of Applied Artificial Intelligence, Sungkyunkwan University. His research focuses on Natural Language Processing, Computer Vision, Deep Learning, and Machine Learning.

**Seon Hong Lee** is PhD student at Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University. His research focuses on Natural Language Processing, Deep Learning, and Machine Learning.

**Jang Hyun Kim** is an professor at Department of Human-Artificial Intelligence Interaction/Department of Interaction Science/Department of Applied Artificial Intelligence, Sungkyunkwan University. His research focuses on social/semantic data analysis, social media, and future media. He has authored over 50 papers in major journals such as Information Processing & Management, Telematics & Informatics, Cities, Government Information Quarterly, Technological Forecasting and Social Change, and Journal of Computer-Mediated Communication. Jang Hyun Kim is the corresponding author and can be contacted via alohakim@skku.edu.