

# A Corpus for Dimensional Sentiment Classification on YouTube Streaming Service

Ching-Wen Hsu  
Department of Big Data Management,  
Soochow University, Taipei, Taiwan  
06170139@gm.scu.edu.tw

Chun-Lin Chou  
Department of Big Data Management,  
Soochow University, Taipei, Taiwan  
07370005@gm.scu.edu.tw

Hsuan Liu  
Department of Big Data Management,  
Soochow University, Taipei, Taiwan  
06170114@gm.scu.edu.tw

Jheng-Long Wu  
Department of Data Science,  
Soochow University, Taipei, Taiwan  
jlwu@gm.scu.edu.tw

## Abstract

The streaming service platform such as YouTube provides a discussion function for audiences worldwide to share comments. YouTubers who upload videos to the YouTube platform want to track the performance of these uploaded videos. However, the present analysis functions of YouTube only provide a few performance indicators such as average view duration, browsing history, variance in audience's demographics, etc., and lack of sentiment analysis on the audience's comments. Therefore, the paper proposes multi-dimensional sentiment indicators such as YouTuber preference, Video preferences, and Excitement level to capture comprehensive sentiment on audience comments for videos and YouTubers. To evaluate the performance of different classifiers, we experiment with deep learning-based, machine learning-based, and BERT-based classifiers to automatically detect three sentiment indicators of an audience's comments. Experimental results indicate that the BERT-based classifier is a better classification model than other classifiers according to F1-score, and the sentiment indicator of Excitement level is quite an improvement. Therefore, the multiple sentiment detection tasks on the video streaming service platform can be solved by the proposed multi-dimensional sentiment indicators accompanied with BERT classifier to gain the best result.

Keywords: Sentiment Analysis, Text Classification, Machine Learning, Deep Learning, Streaming Service

## 1 Introduction

Due to the rapid rise of new media and the popularization of mobile phone networks, audiences' viewing habits have shifted from TV to online social media platforms. Now people can watch videos on different platforms such as Facebook, Dailymotion, and YouTube anytime and

anywhere. YouTube has 16 million active users in Taiwan monthly, and nearly 93% of users have visited YouTube. In addition, YouTube has become ubiquitous and played an increasingly important role in modern life and entertainment. Also, YouTube provides a discussion function for audiences to express their opinion by clicking like or dislike bottom or leaving comments. Therefore, comprehensive sentiment analysis for comments of the audience on YouTube is necessary.

It is verified that public views, comments, and attitudes towards many events can be analyzed through social media (Heredia et al., 2016). Public reviews on Amazon were used to evaluate users' opinions and determine the audience's preference by classifying opinions into negative, positive, and neutral (Bhatt et al., 2015). Another research investigated the popularity of videos by indicators such as the number of likes, dislikes, and views (Chelaru et al., 2013). Social media, especially YouTube, is considered the largest video sharing site, and the platform has developed into a leading marketing tool. (Schwemmer and Ziewiecki, 2018) Inspired by the above analysis tasks and the rapid growth status of YouTube, we propose the multi-dimensional sentiment indicators to analyze comments on YouTube, which aim to help YouTubers check their videos' performance uploaded on the YouTube platform.

In general, sentiment analysis focuses on determining the positive, negative, or neutral emotions in many pieces of research (Cunha et al., 2019). Even if Keith et al. (2016) extend the emotional detection, which includes highly positive, optimistic, neutral, negative, and highly damaging, the emotional variance may have a different dimension, such as excitement which expresses the audience's fluctuating emotion. So,

we also propose detecting the audience's excitement level on YouTube because excitement more precisely determines how much the audience likes Youtubers or videos.

To obtain comprehensive sentiment indicators, we design three indicators: YouTuber preference, Video preference, and Excitement level to analyze a multi-dimensional aspect of the audience's comments. In the experiment, these three sentiment indicators also represent three detection tasks that aim to detect the audience's motivations behind a myriad of comments.

Various models deal with text-based sentiment classification tasks. Machine learning-based models are used to address the text classification task (Sun et al., 2019). Other deep learning models have been used for sentiment analysis and obtained acceptable performances (Hassan & Mahmood, 2017). Recently, it has refreshed the best performance of using pre-trained language models as soon as it appears. ELMo (Peters et al., 2018) and BERT (Radford et al., 2018) have been effective because pre-trained models have learned by detecting other tasks from a larger corpus which capture more linguistic structure.

This paper's objectives are: (1) to create a corpus for multi-dimensional sentiment indicators, which include YouTuber preference, Video preferences, and Excitement level; (2) train an automatic sentiment detection model, including machine learning-based, deep learning-based, and BERT-based models. Overall, the contributions of this paper are: (1) We establish a benchmark dataset of dimensional sentiment classification for analyzing comments on YouTube. (2) Successfully using different models to deal with sentiment classification issues.

## 2 Related Work

More and more researchers undertook experiments on YouTube as the data source. The purpose is to obtain an understanding of the community commenting behavior. Severyn et al. (2016) showed that although most audiences present their opinions as comments, some abuse this mechanism by posting links to external web pages or posting disruptive, false, or offensive comments to fool and provoke other users. Based on the result of the above research, this paper's dataset eliminates non-relative comments such as links that guide people to external web pages and advertisements that have no relation to video content. Schultes et al. (2013)

work on YouTube video comments, likes, and dislikes to show that it genuinely influences users' perceptions of like or dislike towards videos when reviewing valuable comments.

Moving to some purposes of text classification used nowadays, Turney (2002) did sentiment analysis by establishing an unsupervised classifier to judge the positivity or negativity of product reviews (cars, banks, and tourist destinations) and movie reviews. Another paper presented an approach based on a clustering of comment content, leading to appropriate video categories (Leung et al., 2009). Machine learning approaches are then introduced to automatically classify comments according to their usefulness (Bhavitha et al. 2017). As the above papers show, comments can achieve various objectives by using different technical methods.

There are currently two approaches to address sentiment analysis: (1) lexicon-based techniques (2) algorithm-based techniques. Lexicon-based techniques rely on predefined words and rules to guide the sentence towards the tendency of emotion. Algorithm-based techniques can be divided into two groups: machine learning-based models and deep learning-based models.

Zhang and Zheng (2016) discussed machine learning methods for sentiment analysis. Dang et al. (2020) employed deep-learning approaches with word embedding and TF-IDF to solve sentiment analysis problems. As a result, the best behavior when using the word embedding method against TF-IDF of all models has been proved. Another research used pre-trained word embedding as an important component for it downstream models. (T. Miyato, A. M. Dai, et al., 2017) Thus, we identify that word embedding is in conjunction with deep learning-based models in our experimental and pre-trained word embedding offer significant improvement over embedding learned from scratch. Moreover, due to the effectiveness of pre-trained language models, adding one additional output layer can fine-tuned models and accelerating the accuracy of classification problems. Sun et al. (2019) fine-tuned with Bidirectional Encoder Representations from the Transformers (BERT) model and achieved state-of-the-art results using comments. Liat Ein-Dor et al. (2020) using BERT based models for binary classification tasks

To deal with sentiment analysis tasks, these papers all share some commons. Firstly, comments

in nowadays social media, especially YouTube, are of great value and even thoroughly necessary. Secondly, despite different methods that are conducted, comments do reflect users' opinions on social media. The above works are similar to this paper, all using comments as data sources but a different way to solve the task; we reference the above methods and determine to use all the methods, including deep learning-based, machine learning-based, and BERT-based classifiers. However, each method has been experimented with separately so as to carry out a comparative study. Also, the difference is that we focus our experiment on multi-class text classification.

### 3 Methodology

Figure 1 shows the proposed method for sentiment analysis and classification processes as follows: Firstly, we collected the audience's comments from YouTube platform and subsequently labeled these comments to provide meaningful and informative labels such as three sentiment indicators for model training. Data preprocessing works are conducted to clear text. Next, all comment's texts need to be converted into vectors to serve as the model's input. And then, machine learning and deep learning models propose to train detection models for our proposed sentiment indicators. Finally, by the experiment stage, we evaluate the performance of each classifier in three detection tasks and discuss a comparative study.

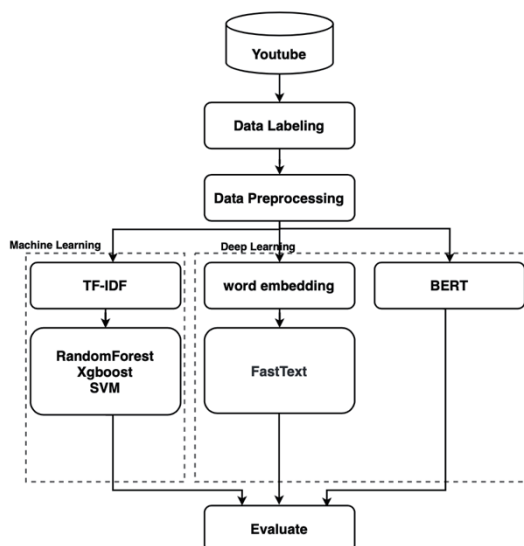


Figure 1: The process of the proposed sentiment analysis in this paper.

### 3.1 Comment Collection

To properly fit data with our analysis targets and cover the diversity of YouTube channels, we select different YouTube channels as our dataset, including 25 YouTuber channels. The composition of the selected videos' film creation types which game with 1%, education 4%, DIY with 4%, science and technology with 5 %, comedy 9%, entertainment with 28%, and blog with 49%. Through these selecting channels, we then filter five videos from each channel that have been highly popular or controversial since 2019 because people imminently show their interest in new trend and debatable topics. Therefore, a total of 25 videos were selected as our data sources. In this way, more controversial and polarizing comments are generated, and it becomes easier to determine the sentimental tendency of comments. However, to avoid different accumulated numbers of comments in each video, we randomly remain 100 pieces of comments from each video. Thus, a total of 12500 pieces comments is taking into consideration.

### 3.2 Definition of Sentiment Indicators for Comment of YouTube

We design three indicators, including YouTube preference, Video preference, and Excitement level, to investigate different aspects of the audience's comment. Each sentiment indicator and the detailed definition is as following:

- **YouTuber preference:** Comments can roughly divide into non-relative and relative towards YouTubers. However, YouTubers may be more concerned about relative comments because these comments help improve YouTubers' behaviors, so we subdivide relative comments into three attitudes towards YouTubers: unlike, neutral, and like. For example, comments not containing YouTuber's name or affair will be labeled as non-relative, and the rest of the labels can determine the audience's tendency of their preferences. Overall, the indicator, YouTuber preference, is categorized as non-relative, unlike, neutral, and like.
- **Video preference:** The indicator, Video preference, is classified into four parts as YouTuber preference. Non-relative, unlike, neutral, and like are four

categories used to judge Video preference. For example, if comments did not contain video content or talk about YouTuber’s affair, then comments are be labeled as non-relative comments. In contrast, comments discussing videos, whether showing their preference, may be labeled as one of unlike, neutral, or like towards video.

- **Excitement level:** The Excitement level, which shows the audience’s emotional ups and downs, is designed into five categories, classifying the audience’s speaking tone from no emotion to extreme emotion state step by step. Moreover, we consider emojis a judgment in this indicator because of the audience’s switching habit in leaving comments. People use a variety of emojis as an emotional expression nowadays, and thus emojis are highly accompanied by texts. Thus, a higher number of emojis containing in comments, a larger Excitement level and sentiment are expressed. For example, the number 0 stands for barely excited emotion contained in comments, while the number 4 represents hyper excited emotion.

### 3.3 Sentiment Indicator Labeling

In this paper, there are three experts to annotate sentiment indicators. All experts possess the background of using YouTube for an extended period and use the YouTube platform frequently. During the annotation process, we eliminate some non-relative comments, such as advertisements, comments that not using Mandarin, comments that post links to external web pages, and merely timestamps in the comments, to optimize the availability of the dataset. Also, to address semantic comprehension gaps between each annotator, we even provide an annotation guideline to consistently label the audience’s comments. Table 1 is a guideline of annotation for the Excitement level indicator. When marking indicator of Excitement Level, sentences with emojis must not be allow to mark as 0 points. Besides, watching the videos is also required before labeling comments; in this way, annotators might resonate powerfully with the audience’s opinions.

Excitement level	Definition
Barely excited	No emoji
Slightly excited	One type of emoji
Excited	Speak confidently and contain two types of emojis
Fairly excited	Emojis are highly repetitive or over three types emojis
Hyper excited	A lot of rhetoric and a series of emojis

Table 1: Annotation guideline to Excitement level.

In Table 2, we show the result of annotation agreement scores using three assessments, including Krippendorff’s Alpha, Fleiss’s Kappa, and Cronbach’s Alpha. With Krippendorff’s Alpha method, due to the reason that values smaller than 0.667 represent as discard data, so our three indicators are shown not up to the standard. Fleiss’s Kappa method stands for fair and moderate data because values between 0.21 to 0.6 are considered acceptable levels. Cronbach’s Alpha method evaluates three indicators as outstanding labeling work because a value higher than 0.7 may show annotation agreement, let alone we get 0.9 on Excitement level. Therefore, two of the methods were qualified as acceptance results, and thus we provide an adequately labeled dataset to train and assess a given model.

	YouTuber preference	Video preference	Excitement level
Krippendorff’s Alpha	0.5829	0.4545	0.3898
Fleiss’s Kappa	0.5840	0.4594	0.3928
Cronbach’s Alpha	0.8520	0.7264	0.900

Table 2: Annotation agreement scores for each indicator.

### 3.4 Text Preprocessing

To deal with a few variances in our annotated results, we use the majority decision to filter out inconsistent labels unless each comment annotation is marked as the same point. This objective is to provide a ‘ground truth,’ a properly labeled dataset, to train and assess a given model. As we mentioned in section 3.3, we consider emojis emotional expressions, so dealing with rich emojis is our priority. We transfer emojis to text by the package called “emojiswitch.” Then, we establish a user-defined dictionary to recognize specific words. For example, we establish the

names of the lead actors/actresses and the supporting actors/actresses from our selected videos. Additionally, texts transferred from emojis are also defined as unique objects and be part of our defined dictionary. In this way, we can increase the accuracy of word tokenization. After executing the above two steps, we use the current state-of-art word tokenization tool created by the Chinese Knowledge and Information Processing (CKIP) Group. This tool is available for dealing with tokenization in Mandarin. Through these processes, every word may contain the same meaning as we do data labeling job.

### 3.5 Text Classification

To verify dimensional sentiment classification that we propose several classifiers to learn and detect sentiment indicators. There are three series classifiers and describe:

- **Machine learning-based classifiers:** RandomForest, Xgboost, and SVM (Amrani et al., 2018) are used as methods for experiments. We transform comments into numerical vectors by using TF-IDF to represent each word related to the entire corpus and serve as inputs to fit models.
- **Deep learning-based classifiers:** We utilize FastText, which is bought with word-embedding in our experiment stage. Because of the lack of a myriad of training Chinese corpus, we take advantage of pre-trained word embeddings based on the 2021 Wikipedia Chinese corpus to transform our data into vectors and use them as inputs to train deep learning-based algorithms. Such a massive corpus may get better feature learning than we train word vectors from our dataset.
- **BERT-based classifier:** Using the pre-trained models (Devlin et al., 2018): We select “distilbert-base-multilingual-cased” and “bert-base-multilingual-cased” as our models. According to the mechanism of pre-trained tokens, the inputs are the output of transferring text using a pre-trained corpus, with 21 thousand words in size. However, not using the word-embedding method as model training,

only adding a unique embedding ([CLS]) before the first word of tokens.

### 3.6 Classification Tasks

We apply three methods, six models, to train classifiers and analyze three targets to capture comprehensive sentiment on the comment of the audience. The following elaborates the meaning of three tasks for our experiment.

- **T1:** The audience’s sentiment towards YouTubers is an extended issue from an indicator of YouTuber preference. We exclude non-relative comments and remain comments of unlike, neutral, and like from the indicator. Like and dislike can serve as a hallmark for YouTubers to check the performance of his or her channel. Also, YouTubers can know what attractive they own or what causes them to make a nuisance.
- **T2:** The audience’s sentiment towards videos excludes non-relative comments from the indicator of Video preference and remains the rest of the comments, including comments of unlike, neutral, and like, just like T1 does. Even if watching the same channel, the different themes will captivate and engage different audiences. Therefore, this task may help YouTubers understand their audience’s preferences within a specific channel.
- **T3:** Corresponding to the indicator of Excitement level, T3 aims to analyze the audience’s emotional ups and downs, which can firmly confirm the degree of support from different audiences and affirm the audience’s attitude towards specific issues.

## 4 Experiment

### 4.1 Dataset

After excluding the non-relative dataset from the indicator of YouTuber preference, most rest comments are labeled as like in the audience’s sentiment towards YouTubers. Next, the composition of comments towards Video preference shows that 60 percent of comments are neutral attitudes. T3 applies the result of the

indicator of Excitement level, revealing that the audience could express their happiness and wrath by commenting. Table 3 shows the proportion of data to our three tasks.

Task	Class	Number
T1	Unlike	287 (10%)
	Neutral	784 (28%)
	Like	1,705 (61%)
T2	Unlike	659 (7%)
	Neutral	5,842 (60%)
	Like	3,274 (33%)
T3	Barely excited	2,788 (30%)
	Slightly excited	2,478 (27%)
	Excited	2,341 (25%)
	Fairly excited	1,136 (12%)
	Hyper excited	471 (5%)

Table 3: Distribution of five tasks.

## 4.2 Experiment design

In this section, we introduce the process of building multiple classifiers. Multiple models are shown in Table 4 and are conducted with different parameters. Through experiments, we configure the best parameters on each model to predict different aspects of sentiment analysis.

Model	Description
M1	BERT model using <i>bert-base-multilingual-cased</i> pre-trained model.
M2	BERT model using <i>distilbert-base-multilingual-cased</i> pre-trained model.
M3	RandomForest + TF-IDF
M4	Xgboost + TF-IDF
M5	SVM + TF-IDF
M6	FastText + embedding

Table 4: There are six models use to solve three tasks.

We use 5-fold cross-validation to ensure the performance for all models. By fixedly set  $k=5$  to our dataset, 80% of the data for training and 20% for testing in each fold. After conducting experiments, we evaluate and interpret the performances of different models through the suitable metrics used for classification problems: overall accuracy and F1-score. These tasks are all be performed by Google Colab GPU.

Selecting the correct parameters is vital to attain maximizing model performance. A set of experimented parameters based on their influence

on the models are conducted in our paper. In deep learning and BERT experiments, parameters such as batch sizes (32 and 64), dropout rates (0.1 and 0.5), and learning rates (0.001 and 0.005) are considered.

By experimenting with numerous combinations of parameters, finally, we configure the best parameter for each algorithm and use it to predict the test dataset. However, the classification in each indicator may not be equally distributed, so accuracy is not efficiently reflecting the model’s performance. Thus, we also use F1-score to measures models’ performance.

## 4.3 Results

Figure 2 shows the result of the audience’s sentiment towards YouTubers. The threshold of model performance is set as 0.5 according to the performance of machine learning-based algorithms. BERT-based classifiers and deep learning-based classifier have similar outcomes, and thus are all better than the machine learning-based classifiers.

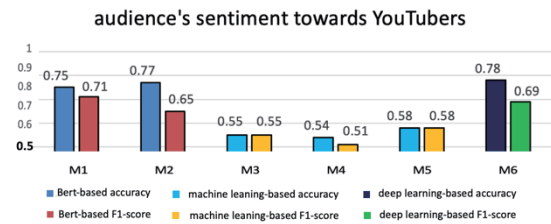


Figure 2: Performance of models on audience’s sentiment towards YouTubers (T1).

Figure 3 is the result of predicting the audience’s sentiment towards videos. We set the threshold of 0.5 according to the performance of machine learning-based algorithms. M3, M4, and M5 achieve the same score in each of their accuracy and F1-score. However, BERT and deep learning-based methods show the same tendency: accuracy is 10% higher than F1-score. It proves that whether models the F1-score of machine learning-based algorithms can highly perform as the accuracy.

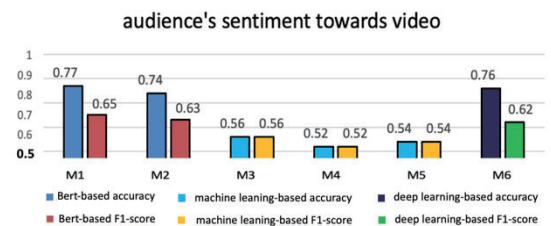


Figure 3: Performance of models on audience’s sentiment towards videos (T2).

As Figure 4, we set the threshold as 0.3 to be the baseline of our models' performance. Compared with the above two tasks, predicting the audience's emotion differs significantly in each method. Nevertheless, this task is relatively the best to distinguish the performance of different methods. For example, scores of machine learning-based classifiers reduce significantly compared with detecting the audience's sentiment towards videos, nearly 20% decrease in accuracy and F1-score. Although BERT and deep learning-based models also drop their performance compared with detecting the audience's sentiment towards videos by 10%, these two methods have the better efficacy of dealing with a multi-classification problem.

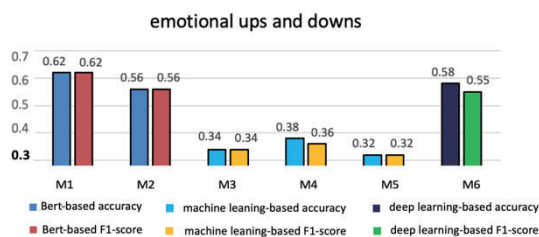


Figure 4: Performance of models on emotional ups and downs (T3).

#### 4.4 Discussion

In summary, three findings follow (1) Within three sentiment detection tasks for comments on YouTube, machine learning-based classifiers cannot achieve performances compared with other methods. (2) When comparing three methods' performance in each detection task, it shows that the performance of the deep learning-based models evenly matched the score with BERT-based models. However, a slight variance exists in F1-score. (3) BERT slightly outperforms other models in three tasks according to the F1-score, and F1-score also achieves its accuracy, which stands for the minority of dataset's categories that are taken into consideration during model predicting.

Nevertheless, the task of the audience's sentiment indicators prediction has solved by this paper, and the BERT model has obtained a significant difference which has 0.62 F1-score improvement over these machine learning-based models. We can also highlight that most comments in the indicator of Excitement level are labeled as barely excited or slightly excited as our training dataset; only a few comments are labeled as having hyper excited. However, few labels obtain nearly the same recall as the majority of labels in our

result. Therefore, the BERT-based models have learned some sentiment patterns from comments of the audience's extreme emotions. Moreover, the experimental result presents that the TF-IDF method has not obtained good performance, because the context of comment is a very important factor but TF-IDF does not handle that.

## 5 Conclusion

This paper focuses on sentiment analysis using the core of BERT pre-trained language models and accompanied by one deep learning-based model and three machine learning-based models. After conducting experiments, the method of deep learning and BERT perform better than the machine learning method. We also show that BERT can deal with sentiment polarity by determining the audience's likes or dislikes towards YouTubers. Finally, BERT is perfectly addressing the multi-classification problem. Before utilizing these classifiers, introducing related labeling jobs as a prerequisite is vital to getting a reliable dataset. Through these methods, we genuinely fill the gaps between human semantic comprehension.

Analyzing the public's perception of YouTubers and the influence of their videos is a challenging task for researchers so far. Proposing different sentiment indicators and utilizing different classifiers has been done in this paper, but there still is a long way to overcome some problems. In this paper, we have emphasized the following problems in order to make our results improve. (1) Informal language styles such as sparse emojis used by the audiences may impede models from capturing linguistic structure. (2) the semantic comprehension gap among annotators needs to be reduced to improve annotation consistency.

In the future, we may explore other techniques for optimizing multiple-dimensional sentiment analysis tasks, such as training YouTubers' names as embedding before utilizing different models. In this way, perhaps models can precisely filter out non-relative comments towards YouTubers. In addition, others indicators, such as whether the comments contain an ironic statement or whether the comments are erotic, can be added for analyzing other aspects of the audience's comments. The latter proposed indicator may serve as a guard for children users, and the former indicator may prevent YouTubers from getting into conflict with their fans.

## References

- Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment analysis using deep learning techniques: a review. *Int J Adv Comput Sci Appl*, 8(6), 424-433.
- Al Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 127, 511-520.
- Bhatt, A., Patel, A., Chheda, H., & Gawande, K. (2015). Amazon Review Classification and Sentiment Analysis. *International Journal of Computer Science and Information Technologies*, 6(6), 5107-5110.
- Bhavitha, B., Rodrigues, A.P., & Chiplunkar, N.N. (2017, March). Comparative study of machine learning techniques in sentimental analysis. In *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 216-221). Coimbatore, India: IEEE.
- Cunha, A.A.L., Costa, M.C., & Pacheco, M.A.C. (2019, June) Sentiment Analysis of YouTube Video Comments Using Deep Neural Networks. In *International Conference on Artificial Intelligence and Soft Computing (ICAISC)* (pp. 561-570). Zakopane, Poland: Springer.
- Dang, N. C., Moreno-García, M. N., & Prieta, F. D. L. (2020). Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*, 9(3), 483-512.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. <https://arxiv.org/abs/1810.04805>
- Keith Norambuena, B., Lettura, E. F., & Villegas, C. M. (2019). Sentiment analysis and opinion mining applied to scientific paper reviews. *Intelligent Data Analysis*, 23(1), 191-214.
- Hassan, A., & Mahmood, A. (2017, April). Deep learning approach for sentiment analysis of short texts. In *2017 3rd international conference on control, automation and robotics (ICCAR)* (pp. 705-710). Nagoya, Japan: IEEE.
- Heredia, B., Khoshgoftaar, T. M., Prusa, J., & Crawford, M. (2016, July). Cross-domain sentiment analysis: An empirical investigation. In *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)* (pp. 160-165). Pittsburgh, PA, United States: IEEE.
- Leung, J. K.-W., Li, C.H., & Ip, T. K. (2009). Commentary-based Video Categorization and Concept Discovery. In *Proceedings of the 2nd ACM workshop on Social web search and mining (SWSM '09)* (pp. 49-56). New York, United States: Association for Computing Machinery.
- Dor, L. E., Halfon, A., Gera, A., Shnarch, E., Dankin, L., Choshen, L., ... & Slonim, N. (2020, November). Active learning for BERT: An empirical study. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 7949-7962).
- Pandey, A.C., Rajpoot, D.S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, 53(4), 764-779.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. <https://arxiv.org/abs/1802.05365>
- Schultes, P., Dorner, V., & Lehner, F. (2013). Leave a Comment! An In-Depth Analysis of User Comments on YouTube. *Wirtschaftsinformatik*, 42, 659-673.
- Schwemmer, C., & Ziewiecki, S. (2018). Social Media Sellout: The Increasing Role of Product Promotion on YouTube. *Social Media + Society*. 4(3). <https://doi.org/10.1177/2056305118786720>
- Severyn, A., Moschitti, A., Uryupina, O., Plank, B., & Filippova, K. (2016). Multi-lingual opinion mining on YouTube. *Information Processing & Management*, 52(1), 46-60.
- Siersdorfer, S., Chelaru, S., Pedro, J. S., Altingovde, I. S., & Nejd, W. (2014). Analyzing and mining comments and comment ratings on the social web. *ACM Transactions on the Web (TWEB)*, 8(3), 1-39.
- Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. <https://arxiv.org/abs/1903.09588>
- T.Miyuto,A. M.Dai, and I. Goodfellow, “Adversarial training methods for semi-supervised text classification,” in The 5th International Conference on Learning Representation (ICLR 2017), 2017.
- Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. <https://arxiv.org/abs/0212032>
- Zhang, X., & Zheng, X. (2016, July). Comparison of text sentiment analysis based on machine learning. In *2016 15th international symposium on parallel and distributed computing (ISPDC)* (pp. 230-233). Fuzhou, China: IEEE.