# Sentiment Analysis Model for KlikIndomaret Android App During Pandemic Using Vader and Transformers NLTK Library

A. G. Budianto[1], B. Wirjodirdjo[2], I. Maflahah[2], D. Kurnianingtyas[3]
[1]Department of Mechanical Engineering, Universitas Lambung Mangkurat, Banjarmasin, Indonesia
[2]Department of Industrial Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[3]Department of Informatics Engineering, Universitas Brawijaya, Malang, Indonesia
(ghiffaryb04@gmail.com, budisantoso.wirjodirdjo@gmail.com, iffanmaflahah@gmail.com, divaku@ub.ac.id)

*Abstract* - **COVID-19 has changed the Indonesian people's shopping habits for consumer goods. The online retail application came as a response to social distancing and stay-at-home advice. KlikIndomaret is an online retail application that uses the omnichannel concept. As the number of downloads increased, the number of various comments and sentiments on that application also increased. In this study, the researcher did a sentiment analysis aimed to improve the quality of application experiences and retail services. The result of the analysis reflected the services given to customers thus far. The data included reviews and star ratings derived from 4,066 reviews which went under the process of data pre-processing. The methods used in this study were VADER and NLTK, improved by Transformer, without pre-training data. These methods could filter the users' reviews with sarcasm tone. The results were sentiment labels that were appropriate based on the score comparison of positive and negative sentiments in one user's review. This approach made the review sentiment process of thousands of data faster and more accurate.**

*Keywords* – **Sentiment Analysis, Transformers, VADER, KlikIndomaret, Text Mining**

## I. INTRODUCTION

Covid-19 has changed the Indonesian people's shopping habits for consumer goods. The advice to stay at home and avoid direct contact became a commonly heard slogan. Technology exists to respond to that challenge. The pandemic era made the retail sectors develop digital and online shopping options [1]. Shopping through omnichannel became the most practical choice for the consumers. The process of omnichannel on the retailed marketplace happens when customers buy goods online and take them directly from the store [2] or buy them online and use the instant delivery service in the pandemic era [3].

The smartphone application, especially the android one, made the shopping process easier. Many e-commerce retailers are available based on the users' demography, such as Amazon and Alibaba [4]. Indomaret is one of the biggest retailers with around 19.000 stores and 38 distribution centers spread in the Indonesia area. Indomaret was a traditional retailer that turned into an omnichannel in the COVID-19 pandemic era [5]. Klikindomaret is an application that has been downloaded more than one million times, has 76.500 ratings and has 5.500 commentary reviews on the Google Play Store [6]. Klikindomaret makes it possible to shop from home, deliver the goods using the instant delivery service, and pick up the goods in the store with a flexible paying system. In improving the quality of service, the ratings and comments on the Google Play Store play a vital role as they reflect the quality of the application and its service.

Sentiment analysis is a technique that studies the perceptions and characteristics of a group to explain the credibility and review of content [7]. Sentiment analysis is related to consumer reviews and information through subjective or objective sentences [8]. Sentiment analysis of consumer opinions on the rating and review system produces positive, neutral, and negative sentiments [9]. Many sentiments analysis is done by using machine learning, text mining, natural language programming (NLP), and classification. In the field of industry and digital business, especially the ratings and reviews on the Google Play Store [9] [10] [11], comments on purchasing goods on Amazon [12] [13] and Alibaba [14] show that sentiment analysis is used to see the users' opinions and reactions of updates that are made to the applications and real-world services. This would certainly help the management in making important decisions regarding omnichannel in the future.

VADER (Valence Aware Dictionary for Sentiment Reasoning) uses qualitative and quantitative methods to generate and validate a contextualized sentiment lexicon. This combination can improve the accuracy of sentiment analysis models in several domains such as social media, film reviews, and product reviews [15]. The use of VADER in Twitter data analysis has proven to be effective, easy to use, and fast in classifying sentiment [16]. Another study related to the sentiment analysis of Bitcoin, a digital object, shows that the tweet sentiment from the VADER model has a short-term correlation with Bitcoin prices [17].

NLP Transformer as a pre-trained method for word and sentence embedding shows good performance for NLP. The task of this transfer learning allows unlabeled data to be used with the small-labeled data to achieve higher accuracy [18]. A Transformer is a deep neural network architecture based on the attention mechanism. It replaces the most used repeating layer with multi-head self-attention [19]. Up to this day, Transformer has been the mainstay architecture for pre-trained models [20]. The Transformer is also used in the detection of sarcasm sentiment comments on Twitter [21].

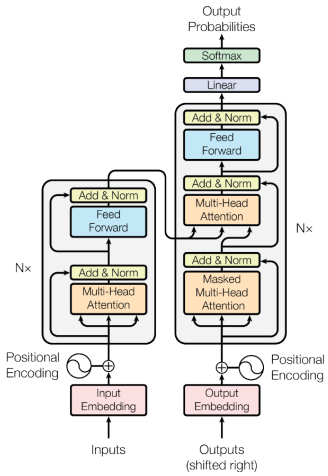


Fig. 1. The Transformer – model architecture [19]

Transformers, combined with deep learning, can improve the tweet quality by removing noise while considering word sentiment, polysemy, syntax, and semantic knowledge [22]. However, sentiment analysis based on machine learning and the Natural Language Toolkit (NLTK) is still constrained by reading subjective data and sarcasm. For example, a customer gives a rating of 5, but the comment section is filled with negative sentiments and vice versa. It's even more complicated with a three-star rating. There are so many interpretations of the comments given, they can have a positive or negative sentiment in a comment. This can be minimized by using the Transformer feature as a sarcasm filter.

Based on the problems described above, this research focuses on making sentiment analysis using VADER that is improved by Transformers on the Klikindomaret android application. The purpose of this study was to reflect on the quality of application and its services during the COVID-19 pandemic in Indonesia and to give some input to the management for their decision-making in the retail omnichannel field. Additional sarcasm filter clarifies the important, but forgotten, input from the results of NLP ordinary sentiment analysis.

## II. METHODOLOGY

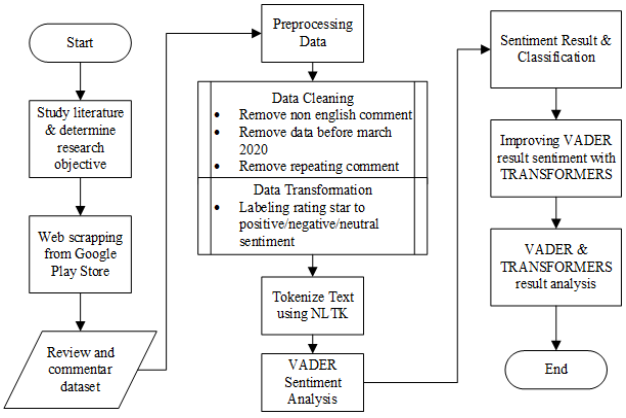


Fig. 2. KlikIndomaret sentiment analysis flowchart

Fig. 2 shows the approach used in this research. It starts by comparing some sentiment analyses approaches such as Naive Bayes, SVM, and Neural Network in study literature. These methods need to conduct data training and depend on the subjective label given by the reviewers of the application. Therefore, by considering the sarcasm and subjectivity of the reviewer, VADER and Transformer methods are considered suitable for sentiment analysis. This is also the goal of the research: obtaining non-fixated sentiment information on the star rating in the Google Play Store application during the pandemic era in Indonesia (March 2020-May 2022)

For data collection, web scraping was the method used with the help of the Python library (Google-Play-Scraper). The token used in this web scraper was 'com.indomaret.klikindomaret'. The data collected were 5,000 review comments with various star ratings. Fig. 3 shows web scrapping process.

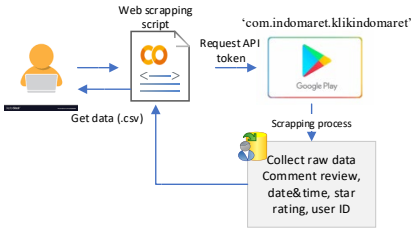


Fig. 3. Web scrapping process

The data were converted to .csv format for further data pre-processing. The datasets are sorted from newest to oldest. Only the data from March 2020 to May 2022 were used. The column of comment data is filtered in English and checked so that no repetition of comments occurs. There were 4,066 rows of data after the completed pre-processing data. Then, the star rating is changed into sentiment labels: negative sentiment is equal to 1-2 stars, the neutral label is equal to 3 stars, and positive sentiment is equal to 4-5 stars. This aimed to form structured data that depend on the subjectivity of users and star ratings.

Dataset processing was continued by using NLTK. The thousands of comments were processed into per word tokens to facilitate the polarity of reviewers' comments. Then, VADER was used to give negative, neutral, positive, and compound scores for each comment. The results on VADER were used to see the sentiments based on the compound value.

After the VADER approach gave a compound value to each comment, the Transformer method was used to improve the quality of the compound and evaluate the final positive/negative label. In the end, an experiment was conducted involving sentences with sarcasm tone from the KlikIndomaret application. The results of this model are discussed further in this paper.

## III. RESULTS

The statistics of data collected from the Google Play Store KlikIndomaret are shown as follows in Fig 4.

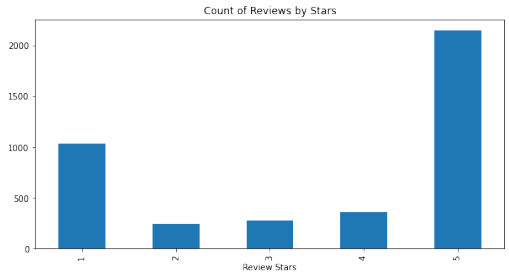


Fig. 4. The star rating by the KlikIndomaret application users

After we use NLTK and VADER sentiment analysis. The result gives us a compound score for each comment review shown in Fig. 5.

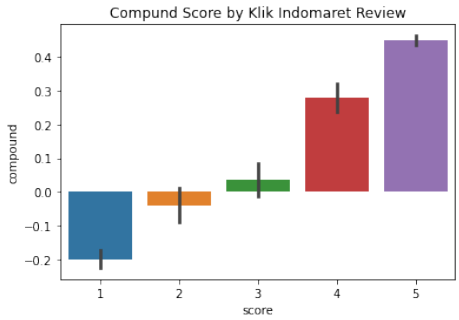


Fig. 5. Compound score results from the VADER method for all the star ratings

The bar plot assessment for positive compounds was considered positive, the neutral one was neutral, and the negative one was negative. This bar plot shows that sarcasm comments exist in the comments and star ratings of this application. Especially on the neutral label (3 star-rating) shown in Fig. 6. So many comments are labeled neutral but on the other hand it has so many interpretations. It can be positive or negative sentiment. But if we look to bar plot, mostly to a negative sentiment (neutral label with score 1-2 have 0.8 compound score).

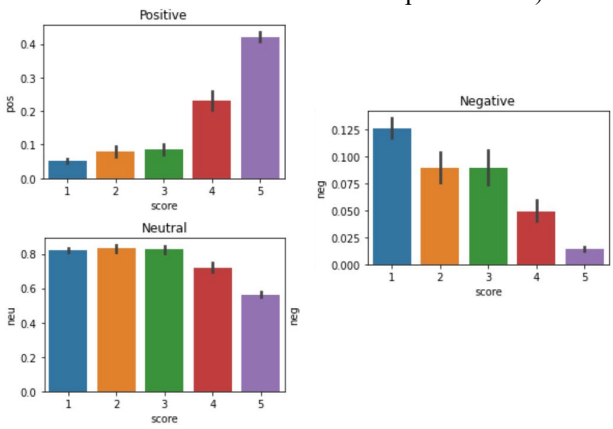


Fig. 6. Bar plot for positive, neutral, and negative compound results based on the star rating

After the NLTK and VADER method improved by TRANSFORMER was processed. It takes 8 min 51 sec to processed 4.066 comments review, the pair plot for sarcasm comment filter was generated in Fig. 7 as follows.

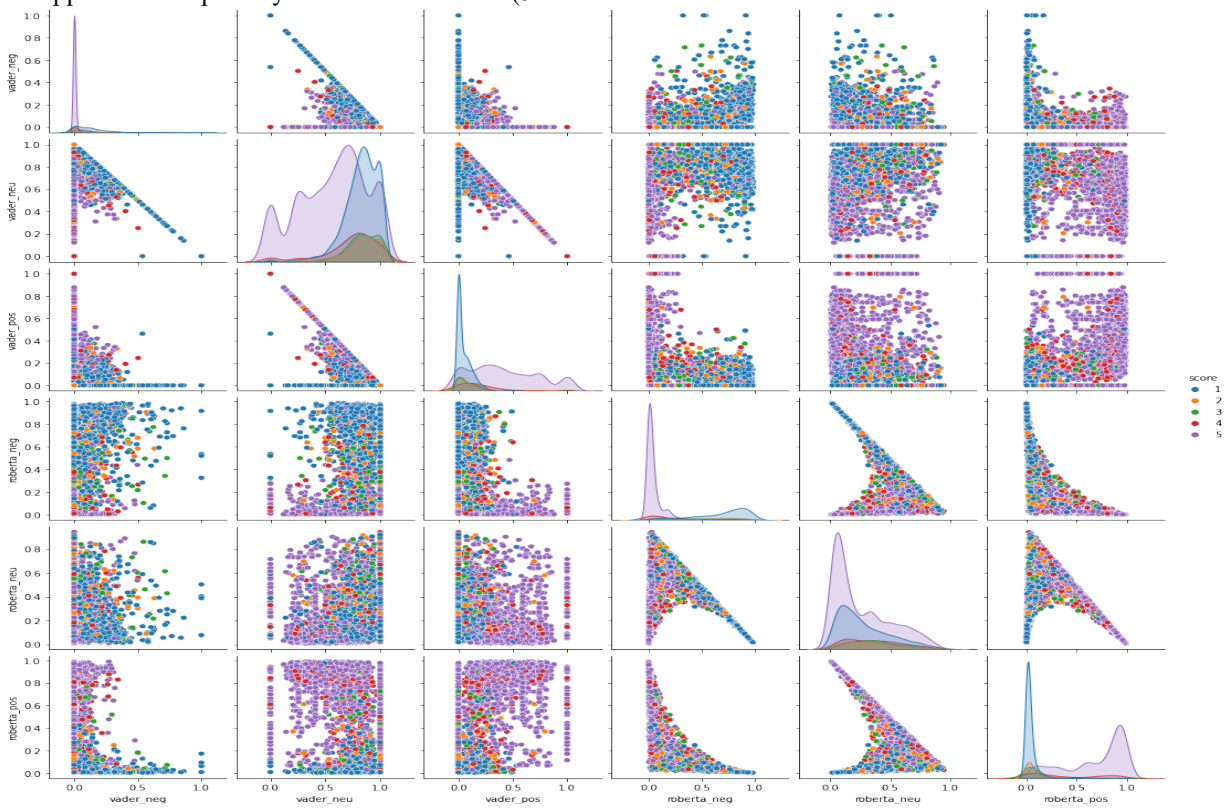


Fig. 7. The distribution of paired data, which were considered positive by VADER, but considered negative by Transformer and vice versa

Fig. 8 show us an example of VADER improved by Transformers from KlikIndomaret reviews

```
[ ] results_df.query('score == 1') \
        .sort_values('roberta_pos', ascending=False)['review'].values[0]

    'from the application is really good, hopefully ranked 1'

[ ] results_df.query('score == 1') \
        .sort_values('vader_pos', ascending=False)['review'].values[0]

    'Stock not updated. Bad.'

[ ] results_df.query('score == 5') \
        .sort_values('roberta_neg', ascending=False)['review'].values[0]

    'Groceries are not appropriate, nothing can be contacted.'

[ ] results_df.query('score == 5') \
        .sort_values('vader_neg', ascending=False)['review'].values[0]

    'Fast response to complaints'
```

Fig. 8. The results of VADER improved by Transformers from KlikIndomaret reviews

After the NLTK and VADER model was improved by Transformer, the researcher tried it on three sentences from comments that have positive and negative sentiments in the same sentence. The result is shown in Figure 9.

```
[33] sent_pipeline('I love shopping using klikindomaret,
     'if the price is competitive with another online retail ')

     [{'label': 'POSITIVE', 'score': 0.6294752359390259}]

[34] sent_pipeline('i try to use this app for the first time,
     'it is simple to use but sometimes the app crash when i use it')

     [{'label': 'NEGATIVE', 'score': 0.9995577931404114}]

[35] sent_pipeline('this app simple and helpful when pandemic.
     'I use the delivery order service but the item i want to buy always out of stock')

     [{'label': 'NEGATIVE', 'score': 0.7609668374061584}]
```

Fig. 9. Prediction results of sentiment labels and scores from VADER improved by Transformer without pre-trained data

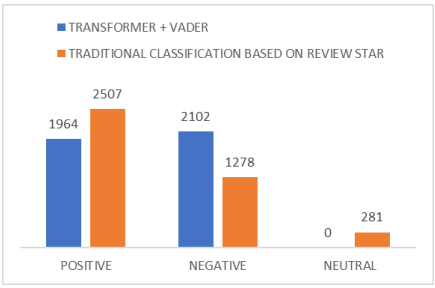


Fig. 10. Sentiment label count summary from 4.066 comments

Fig. 10 shows that the VADER & Transformer model can filter user review comments. It is proven that many users initially gave a positive rating, but the comments contained complaints. And for the neutral comments can be classified well into one of the sentiment labels.

## IV. DISCUSSION

The NLTK and VADER method assigns a compound score based on positive/negative sentiment in a comment. Fig. 5 and 6 show that there are many comments in the review section that give a 5-star rating yet have a negative compound and vice versa. This also applies to the 3-star rating which should be neutral but is divided between positive and negative sentiments. Especially if we look Fig. 5., on neutral label which have low rating (1-2 star) score with a high compound score mostly it's a negative sentiment comment from KlikIndomaret user.

After NLTK and VADER improved by TRANSFORMER, it is proven to be able to filter sarcasm, the inappropriate star rating and comments for app user. It takes 8 min 59 sec to process 4.066 comments review to be labeled as a positive or negative sentiment. This can be seen in Figure 7 which is explained in the following table 1.

TABLE I
The results of the classification of VADER and TRANSFORMER sentiments on review data of KlikIndomaret

| Rating | Result | Comment | Interpretation |
|---|---|---|---|
| 1 | roberta_pos | from the application is really good, hopefully ranked 1 | The user gave a low rating, but Transformer gave positive sentiment. |
| 1 | VADER_pos | Stock not updated. Bad. | The user gave a low rating, but VADER gave positive sentiments. |
| 5 | roberta_neg | Groceries are not appropriate, nothing can be contacted. | The user gave a high rating, but Transformer gave negative sentiment. |
| 5 | VADER_neg | Fast response to complaints | The user gave a high rating, but VADER gave negative sentiment. |

Fig. 9 shows the sentiment analysis model from VADER and NLTK which has been improved by Transformer. It can provide the appropriate sentiment label even though there are positive and negative sentiments in one sentence.


(A) Positive Word Cloud (B) Negarive Word Cloud

Fig. 11. Word cloud based on sentiment label

Fig. 11 shows from positive word cloud, the users praised the KlikIndomaret android app such as ease of use (how to order), and ease of payment, good service & delivery. On the other hand, the users also complained about the response to their complaints (payment trouble and refund), the unavailability of the goods they wanted to buy, time response to deliver the goods, and the

specific address or the contact number to be contacted. Such sentiments should reflect the management in developing its services, especially in the online retail sector and its development in omnichannel.

## V. CONCLUSION

In this study, VADER and NLTK were applied to sentiment analysis of comments and star ratings on the KlikIndomaret application on the Google Play Store without doing the pre-training. The results show the appropriate sentiments between the rating label and the comments given. However, there are some sarcasm comments, such as giving high ratings with negative comments and vice versa, which fail to be classified in the appropriate type of sentiment. The quality of the sentiment can be improved by using a Transformer. The results of the VADER model improved with Transformer can provide a more accurate label of sentiment even though the sarcasm-inspired comment has a composition of positive and negative comments in one sentence. The VADER and Transformer are also proven to classify appropriate sentiment labels of 4,066 comments properly, quickly, and accurately. The process took 8 min 59 sec to give a sentiment label from 4.066 comments. 1.964 user comments were labeled with positive sentiment and 2.102 were labeled with negative sentiment. The positive word that often mentioned are good and helpful app service, order and account payment. On the other hand, negative word that often mentioned are response time, delivery, order, stock, goods, payment and refund

## REFERENCES


[1] S. Ying, S. Sindakis, S. Aggarwal, C. Chen, and J. Su, "Managing big data in the retail industry of Singapore: Examining the impact on customer satisfaction and organizational performance," Eur. Manag. J., vol. 39, no. 3, pp. 390–400, 2021.

[2] F. Gao and X. Su, "Omnichannel retail operations with buy-online-and-pick-up-in-store," Manage. Sci., vol. 63, no. 8, pp. 2478–2492, 2017.

[3] M. Hänninen, S. K. Kwan, and L. Mitronen, "From the store to omnichannel retail: looking back over three decades of research," Int. Rev. Retail. Distrib. Consum. Res., vol. 31, no. 1, pp. 1–35, 2021.

[4] S. Helm, S. H. Kim, and S. Van Riper, "Navigating the 'retail apocalypse': A framework of consumer evaluations of the new retail landscape," J. Retail. Consum. Serv., vol. 54, p. 101683, 2020.

[5] D. Jatimoyo, F. Rohman, and A. Djazuli, "The effect of perceived ease of use on continuance intention through perceived usefulness and trust: A study on Klikindomaret service users in Malang City," Int. J. Res. Bus. Soc. Sci., vol. 10, no. 4, pp. 430–437, 2021.

[6] "KlikIndomaret on Google Play Store." [Online]. Available: https://play.google.com/store/apps/details?id=com.indomaret.klikindomaret&hl=en_US&gl=US. [Accessed: 05-May-2022].

[7] F. N. Ribeiro, M. Araújo, P. Gonçalves, M. A. Gonçalves, and F. Benevenuto, "Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods," EPJ Data Sci., vol. 5, no. 1, pp. 1–29, 2016.

[8] S. Becken, B. Stantic, J. Chen, A. R. Alaei, and R. M. Connolly, "Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing," J. Environ. Manage., vol. 203, pp. 87–97, 2017.

[9] S. W. Handani, D. I. S. Saputra, R. M. Arino, and G. F. A. Ramadhan, "Sentiment Analysis for Go-Jek on Google Play Store," in Journal of Physics: Conference Series, 2019, vol. 1196, no. 1, p. 12032.

[10] S. Fransiska, R. Rianto, and A. I. Gufroni, "Sentiment Analysis Provider by. U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," Sci. J. Informatics, vol. 7, no. 2, pp. 203–212, 2020.

[11] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm," in Journal of Physics: Conference Series, 2020, vol. 1641, no. 1, p. 12043.

[12] K. S. Srujan, S. S. Nikhil, H. Raghav Rao, K. Karthik, B. S. Harish, and H. M. Keerthi Kumar, "Classification of amazon book reviews based on sentiment analysis," in Information Systems Design and Intelligent Applications, Springer, 2018, pp. 401–411.

[13] N. Nandal, R. Tanwar, and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products," Spat. Inf. Res., vol. 28, no. 5, pp. 601–607, 2020.

[14] L. Zhang, S. Fu, and B. Li, "Research on stock price forecast based on news sentiment analysis—A case study of alibaba," in International Conference on Computational Science, 2018, pp. 429–442.

[15] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in Proceedings of the international AAAI conference on web and social media, 2014, vol. 8, no. 1, pp. 216–225.

[16] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and VADER sentiment," in Proceedings of the international multiconference of engineers and computer scientists, 2019, vol. 122, p. 16.

[17] T. Pano and R. Kashef, "A complete VADER-based sentiment analysis of bitcoin (BTC) tweets during the era of COVID-19," Big Data Cogn. Comput., vol. 4, no. 4, p. 33, 2020.

[18] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, "Evaluation of sentiment analysis in finance: from lexicons to transformers," IEEE access, vol. 8, pp. 131662–131682, 2020.

[19] A. Vaswani et al., "Attention is all you need," Adv. Neural Inf. Process. Syst., vol. 30, 2017.

[20] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," Sci. China Technol. Sci., vol. 63, no. 10, pp. 1872–1897, 2020.

[21] S. Javdan and B. Minaei-Bidgoli, "Applying transformers and aspect-based sentiment analysis approaches on sarcasm detection," in Proceedings of the Second Workshop on Figurative Language Processing, 2020, pp. 67–71.

[22] U. Naseem, I. Razzak, K. Musial, and M. Imran, "Transformer based deep intelligent contextual embedding for twitter sentiment analysis," Futur. Gener. Comput. Syst., vol. 113, pp. 58–69, 2020.