

Result Analysis on Twitter Sentiment Analysis using BERT Algorithm

Samruddhi Manoj Mahalle¹, Dr.V.S. Gulhane², and Dr. Avinash P. Jadhao³

¹*Scholar, DRGIT&R, Amravati*

Professor, Information Technology Department, Sipna's CET, Amravati

Associate Professor, Department of Computer Science and Engineering, DRGIT&R, Amravati Institute

Abstract — With the exponential growth of web technology and the increasing popularity of social networking sites like Twitter, there has been an unprecedented volume of user-generated data available on the internet. The diverse nature of opinions and sentiments expressed in tweets presents a unique opportunity for sentiment analysis, where understanding the emotions and attitudes of users towards various topics becomes crucial. Sentiment analysis on Twitter is a challenging task due to the unstructured and heterogeneous nature of the content, which can be positive, negative, or neutral in different contexts. In this paper, we present a comprehensive survey and comparative analysis of existing techniques for sentiment analysis, with a focus on utilizing the BERT (Bidirectional Encoder Representations from Transformers) model, a cutting-edge deep learning approach.

Index Terms— BERT, Sentiment, Twitter, Django, Deep Learning.

I. INTRODUCTION

There has been a massive increase in the usage of microblogging sites like Twitter in the past few years. Businesses and media organizations are increasingly trying to figure out ways to explore Twitter for knowledge about what individuals think and feel about their goods and services, driven by that development. Companies such as Twitratr, tweetfeel, and Social Note are only a few that as one of their offerings promote Twitter sentiment analysis. There exist a decent content of research for identifying the various sentiments that are communicated across various domains like online reviews and news stories. Yet, a few contents of research has been conducted for identifying the various sentiments are communicated considering the informal speech and message-length restrictions of online media. Apps like

automated element tags and techniques like sentimental lexicons have proven effective in many other fields for sentiment analysis, and yet can they still prove beneficial on Twitter for sentiment analysis? The enormous range of subject matter discussed is yet another microblogging issue. Saying that user's tweet on everything and anything is not a misconception. Therefore, one needs a framework for easily defining details that could be utilized for learning is required to be allowed to develop a framework for mining Twitter's feelings on any given subject. Thus, this report discusses one strategy for constructing such information utilizing hashtags that are utilized in Twitter to classify various sentiments like positive, negative, and neutral and utilize various-way emotion classifiers for learning.

Sentiment analysis is the automated mechanism for interpreting a written or spoken thought about a certain topic. Sentiment analysis has been a crucial method for interpreting this data in an environment where daily we produce 2, 5-quintillion bytes of data. It helps businesses to obtain crucial knowledge and optimize operations of all sorts [1].

Sentiment Analysis is often referred to as be Opinion Mining. This builds structures to classify and collect views inside documents. The research is a field of natural language processing (NLP) [1].

Networking sites offer an incentive for companies by offering a forum for ads to communicate with their clients. For decision-making, users often rely to a large degree on customer-based content online. For example, if some user wants to purchase an item or wishes to utilize some utility, the user would be looking for the corresponding ratings online, which follows chat about that on networking sites before making a decision. For a typical user to examine, the

volume of information created by users is too large. There is a big urge to simplify this since different types of sentiment analysis are commonly utilized. Analyzing various sentiments advises the customer whether the goods or the services obtained would be satisfactory not in advance of buying those products. This type of research information is used by advertisers and companies to recognize their goods or services according to the needs of customers that can be delivered.

Sentiment analyses are actually a matter of significant concern because they have a range of functional applications. Because the number of texts that convey views is continuously growing on the Internet, which is publicly and privately accessible, are published on journaling pages, forums, blogs, and social networking [1].

Sentiment Analysis is the automatic method of interpretation and sorting of text data into optimistic, negative, or neutral feelings. Doing a Twitter Sentiment Study using machine learning will help companies realize how customers think about their products [2].

About 80 percent of digital data globally is unstructured and is no exception to data gathered from social network outlets. As the knowledge is not pre-defined, sorting and evaluating are hard. Thanks to Machine Learning and NLP, we are fortunate that models can now be developed, learning from examples and used for processing and organizing text data [2].

Sentiment analytical systems from Twitter can be used to arrange broad tweets and automatically determine the polarity of each message. And, the best thing, is to save precious hours for staff quick and easy and to concentrate on activities that will have a larger impact.

Twitter helps companies to enter a wide market and communicate with consumers without any intermediaries and more than 321 million active users posting a regular total of 500 million Messages. At the drawback, marketers can consider controversial news more easily because if it goes viral, they may wind up facing an unwanted Marketing problem. This is one of the factors that social listening – the monitoring of social network messages and input – has been a critical mechanism for promoting social networks [2].

Monitoring Twitter allows companies to understand their audience, keep on top of what's being said about their brand and their competitors, and discover new

trends in the industry. Are users talking positively or negatively about a product? Well, that's exactly what sentiment analysis determines [2].

Twitter monitoring helps businesses to consider their customers, keep current with their products and competition, and identify emerging market developments. Are consumers concerned about a commodity as good or negative? Yeah, that's what the study of emotions decides [2].

II PROPOSED METHODOLOGY

Proposed Methodology of Twitter Sentiment Analysis using BERT has following stages

Data Collection: The first step in any sentiment analysis project is to gather relevant data. In this case, a dataset of Twitter tweets is required, where each tweet is labeled with its corresponding sentiment (positive, negative, or neutral). There are publicly available datasets specifically designed for sentiment analysis tasks, or you can create a custom dataset by manually annotating tweets with sentiment labels.

Data Preprocessing: Twitter data often contains noise in the form of hashtags, mentions, URLs, and special characters. Preprocessing is crucial to remove these elements and ensure that the data is cleaned and ready for analysis. Additionally, tokenization is applied to split the tweets into individual words or sub words for further processing.

Fine-tuning BERT Model: BERT, a transformer-based language model, has shown exceptional performance in various NLP tasks, including sentiment analysis. However, pre-trained BERT models are not directly suitable for sentiment classification on Twitter data. Therefore, the BERT model needs to be fine-tuned using the labeled Twitter dataset.

Tokenization: It is a critical step to convert the preprocessed text into tokens that the BERT model can understand. Each token is associated with a unique numerical representation, and the tokens are combined to form input sequences for the model.

Building the Classification Model: After fine-tuning the BERT model, it is combined with additional layers to create a sentiment classification model. Typically, a few dense layers are added on top of BERT to adapt it to the specific sentiment analysis task. The last layer of the model is a softmax layer that provides probabilities for each sentiment class.

Training the Model: The model is trained using the labeled Twitter dataset. During training, the model learns to map input tweet sequences to their respective sentiment classes by minimizing a suitable loss function, such as cross-entropy loss.

Model Evaluation: After training, the model's performance is evaluated using a separate validation dataset. Common evaluation metrics include accuracy, precision, recall, F1-score, and AUC-ROC. The goal is to assess how well the model generalizes to new, unseen data.

Hyperparameter Tuning: Hyperparameters, such as learning rate, batch size, and number of training epochs, play a crucial role in the performance of the model. Iterative hyperparameter tuning is performed to optimize the model's performance on the validation dataset.

Sentiment Analysis on New Tweets: Once the model is trained and evaluated, it can be used for sentiment analysis on new, unlabeled tweets. The preprocessing steps applied during training are also applied to the new data, and the fine-tuned BERT model predicts the sentiment for each tweet.

Interpretation and Application: Finally, the sentiment analysis results can be interpreted and applied for various purposes, such as understanding public opinion, tracking sentiment trends, brand reputation management, and real-time monitoring during events. Nevertheless, the utilization of BERT in sentiment analysis has demonstrated promising results and has become a widely adopted approach in the NLP community

III SYSTEM IMPLEMENTATION

The complete implementation of the proposed framework can be implemented in various steps as mentioned in figure-1. These steps were mentioned as follows:

1. Gathered Data:
2. Data pre-processing:
 - a. Tokenization:
 - b. Stop Words Removal:
 - c. Stemming and Lemming:
3. Extraction of features:
4. Identifying the keywords and classifying the movie ratings according to the recognized sentiments.
5. The training dataset and testing dataset need to pass into the proposed model and evaluate the model.

6. Now, the validation dataset will be passed on to the trained and tested model to attain the prediction or identification of polarized sentiments as positive or negative.

Stages of Work Flow in proposed system

Login Page

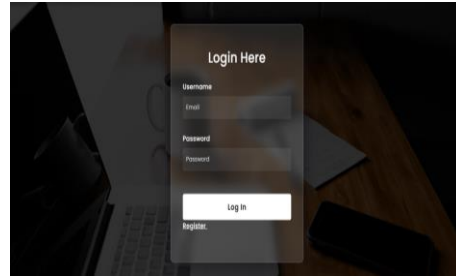


Fig 1 Login Page

Once the user registers themselves, they enter the login panel. The Login panel takes two parameter that is username and password credential from user. Once the password is matched it takes to home page of user.

Login Panel Takes into two dashboard that is user and admin.

User home page consist of various features Admin dashboard has privilege to access the user data

Registration Page

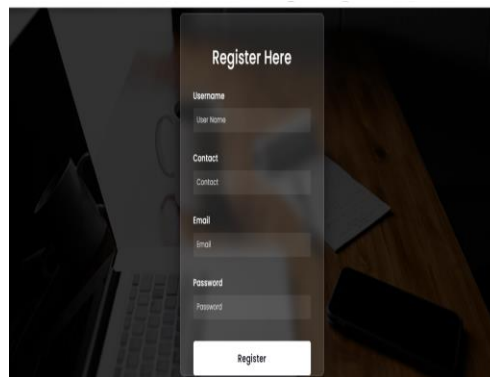


Fig 2 Registration Page

The user who does not have login id have to register themselves to get accessed through application. The registration page consists of various information like name of the person, contact number, email and password of the login. Once the registration is successfully then only, he can login. Once the registration is done registration successful message is displayed.

The email ID and password are required for login.

Home Page



Fig 3 Home Page

The home page module of the user panel in a twitter sentiment analysis can provide a centralized and user-friendly interface for users to access relevant information and features. The home page module in a user-centric manner, keeping the interface intuitive, visually appealing, and responsive across different devices and screen sizes. Conducting user testing and gathering feedback during the development process can help refine the design and ensure it meets the users' needs effectively.

About Us Module

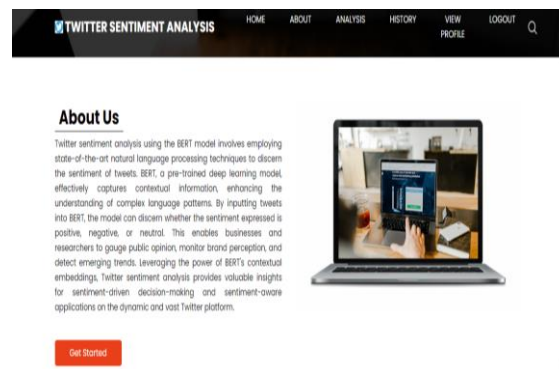


Fig 4 About Us Module

The "About Us" module in the user panel of a twitter sentiment analysis provides information about the system, its purpose, and the team or organization behind it. Here are some key components to include in the "About Us" module:

Provide a brief overview of the stress detection system, explaining its purpose, goals, and how it can benefit users. Describe the main features and functionalities of the system, highlighting how it can help users monitor and manage their stress levels effectively.

Clearly articulate the mission and vision of the organization or team behind the stress detection

system. Explain the driving force behind the development of the system and the broader goals or social impact the team aims to achieve.

Input Text Module

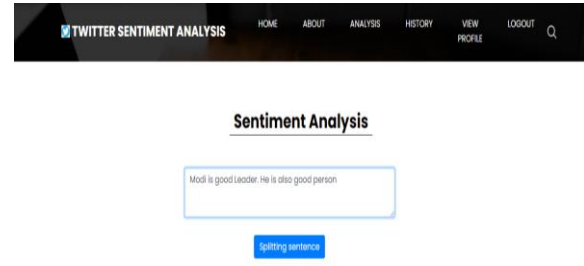


Fig 5 Input Text Module

Our stress detection system utilizes cutting-edge natural language processing and sentiment analysis algorithms to analyze tweets and other textual data from Twitter users. By analyzing the language patterns and sentiments expressed in tweets, we can identify indicators of stress and emotional distress. This enables us to provide valuable insights and support to users who may be experiencing heightened stress levels.

Splitting Sentence Module

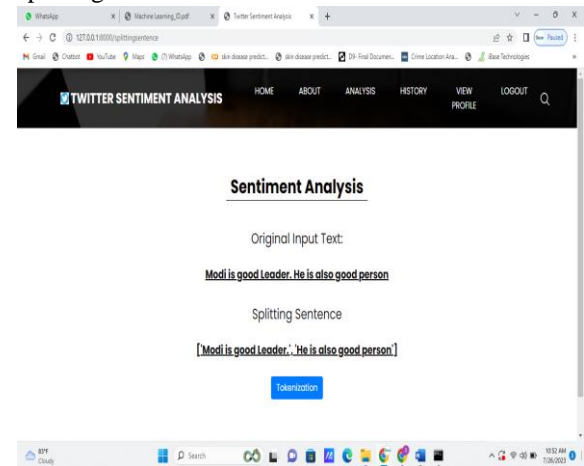


Fig 6 Splitting Sentence Module

The "Splitting Sentence Module" refers to a component or function that is designed to break down a given text or paragraph into individual sentences. The purpose of this module is to facilitate further analysis and processing of the text at a more granular level.

Tokenization Module



Fig 7: Tokenization Module

Internally, the module utilizes a technique called "sentence tokenization" to split the text into sentences. Tokenization is the process of dividing a text into smaller units, such as words or sentences. In our case, we focus on sentence tokenization. It involves identifying sentence boundaries based on punctuation marks (e.g., periods, question marks, exclamation marks) and other linguistic rules. The module leverages existing libraries or custom algorithms to perform this tokenization effectively.

Stop Word Removal Module

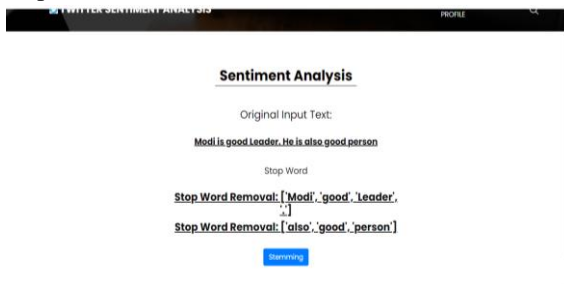


Fig 8: Stop Word Removal Module

The "Stop Word Removal Module" is a crucial part of the preprocessing pipeline in natural language processing (NLP) and text analysis tasks. Its primary purpose is to eliminate common words, known as stop words, from the input text. Stop words are words that appear frequently in a language but do not carry significant meaning and are often used to form the grammatical structure of sentences. Examples of stop words in English include "the," "is," "and," "in," "a," "to," etc.

Sentiment Analysis Module



Fig 9 Sentiment Analysis Module

The Sentiment Analysis Module is a crucial component of our Twitter Sentiment Analysis project, which aims to analyze the emotions and sentiments expressed in tweets from Twitter users. This module plays a pivotal role in extracting insights from textual data, enabling us to understand the overall sentiment of tweets and providing valuable information for various applications.

View History Module

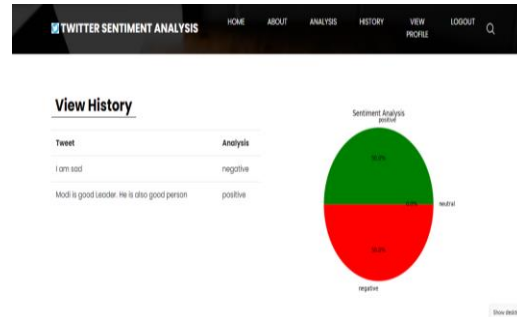


Fig 10 View History Module

The "View History" module is an important component of our Twitter Sentiment Analysis project. This module allows users to access and review their past sentiment analysis results and view the historical sentiment trends related to specific hashtags, events, or their own tweets. The View History module enhances the user experience by providing a comprehensive overview of their engagement and emotional patterns on Twitter over time.

My Profile Module



Fig 11 : My Profile Module

The "My Profile" module in the user panel of a stress detection system provides users with the ability to view and manage their personal information and preferences. "My Profile" module with a user-friendly interface that is easy to navigate and understand. Provide clear instructions and guidance throughout the profile management process to assist users in updating their information

effectively. Regularly update and enhance the module based on user feedback to improve the overall user experience.

Admin Dashboard Module



Fig 12 :Admin Dashboard Module

The admin dashboard module in a twitter sentiment analysis serves as a centralized interface for administrators or system managers to monitor and manage the system's functionality and user data. The admin dashboard module with a user-friendly interface, allowing administrators to navigate and access relevant information easily. Incorporate data visualization techniques to present information in a clear and concise manner, enabling administrators to make informed decisions and effectively manage the stress detection system.

View User Profile Module

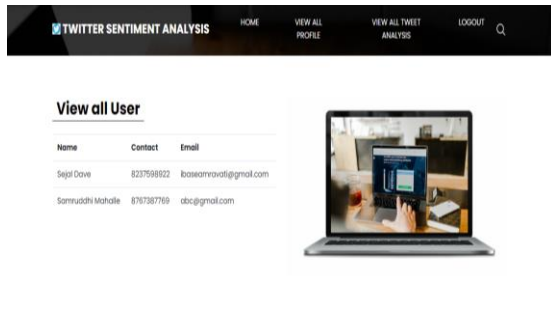


Fig 13 View User Profile Module

The "View User Profile" module allows administrators to access and view detailed information about individual user profiles. It provides a comprehensive overview of the user's account and activity within the stress detection system.

Administrators can access and review specific details related to the user's account settings. This may include information such as account status (active or inactive), role or permissions assigned to

the user, and preferences related to notifications or system configurations.

View All Analysis Module



Fig 14 View All Analysis Module

The "View All Analysis" module is a significant component of our Twitter Sentiment Analysis project, providing users with a comprehensive and centralized view of sentiment analysis results for a wide range of tweets. This module is designed to cater to users who seek a broader understanding of sentiment trends, want to explore sentiment patterns across various hashtags, or analyze sentiments related to specific events or topics.

IV RESULT ANALYSIS & APPLICATION TESTING

Result analysis of Twitter sentiment analysis using the BERT model is aimed at evaluating the performance and effectiveness of the sentiment classification model. It involves measuring how well the model predicts sentiments (positive, negative, or neutral) for the given tweets.

The model's generalization ability is of paramount importance. Ensuring that the BERT model can effectively handle new, unseen Twitter data is vital for its practical application. The generalization assessment allows researchers to ascertain whether the model can be deployed in real-world scenarios and reliably analyze sentiments across diverse Twitter datasets.

By following this data flow and comprehensive analysis, researchers and practitioners can develop robust sentiment analysis models using the BERT architecture, providing valuable insights into public opinion, sentiment trends, and brand perception in the dynamic world of Twitter

Training loss, also known as training error or training cost, is a measure of how well the model is fitting the

training data. During the training phase, the model learns from the training data and tries to minimize the difference between its predictions and the actual target labels. The training loss quantifies this difference. Common loss functions used for training loss include mean squared error (MSE) for regression problems and cross-entropy loss for classification problems.

As the model undergoes multiple iterations (epochs) of training, the training loss ideally decreases. This indicates that the model is improving its ability to capture patterns in the training data and is becoming more accurate in predicting the target labels.

The training and testing loss should decrease and converge to a similar value as the model learns from the data. However, if the testing loss starts to increase while the training loss keeps decreasing, it indicates overfitting, and regularization techniques may be necessary to improve generalization

Parameters	Values
Accuracy:	83.27931428485229
Precision:	95.21
F1 Score:	80.2

Table 1: Values for result analysis

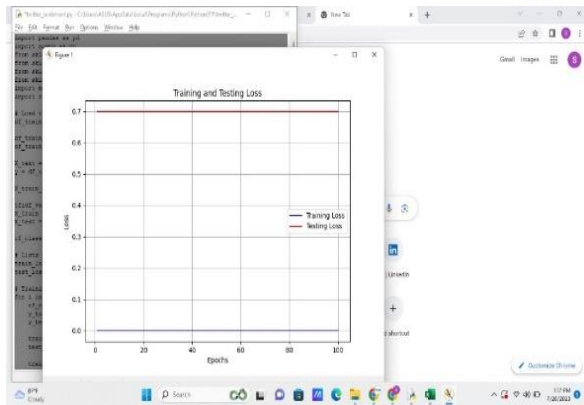


Fig 15 Obtained results

The proposed framework was implemented and attained the confusion matrix as well as the accuracy, precision, recall, and F1-score. The confusion matrix can be represented as mentioned in figure-2. Depending on this confusion matrix, the accuracy, precision, recall, and F1-score can be evaluated and can be compared with the obtained values. The various terms that are used in the confusion matrix can be defined as follows:

1. TP: The full form of TP is true positive which refers to the number of cases that are classified as positive in the actual scenario also represents positive.
2. FN: The full form of FN is false negative which refers to the number of cases that are classified as positive in the actual scenario represents negative.
3. FP: The full form of FP is false positive which refers to the number of cases that are classified as negative in the actual scenario represents positive.
4. TN: The full form of TN is true negative which refers to the number of cases that are classified as negative in the actual scenario also represents negative.

From this confusion matrix, we can evaluate the various metrics as mentioned as follows:

$$1. Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$2. Precision = \frac{TP}{TP + FP} \quad (2)$$

$$3. Recall = \frac{TP}{TP + FN} \quad (3)$$

$$4. F1 - score = \frac{2 * Recall * precision}{Recall + Precision} \quad (4)$$

The obtained confusion matrix can be represented as mentioned in figure with consideration of 1000 instances for the validation aspects.

The graph is color-coded, with different shades of blue used to indicate the number of instances for each combination of true and predicted labels. The color intensity reflects the count of data points falling into each category.

The diagonal elements of the confusion matrix represent correct predictions (true positives and true negatives). For example, the top-left cell (row=0, column=0) shows the count of instances where the true label and the predicted label are both in class 0 (True Negative). The bottom-right cell (row=n, column=n) shows the count of instances where the true label and the predicted label are both in class n (True Positive). The off-diagonal elements represent incorrect predictions (false positives and false negatives). For example, the cell (row=0, column=1) shows the count of instances where the true label is in class 1, but the model predicted it as class 0 (False Negative). The cell (row=1, column=0) shows the count of instances where the true label is in class 0, but the model predicted it as class 1 (False Positive).

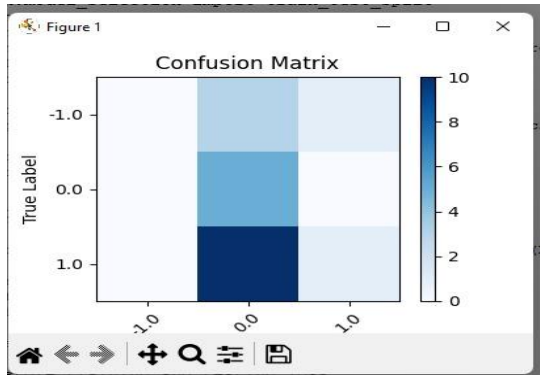


Fig 16 Obtained confusion matrix from the proposed framework

The obtained evaluation metrics obtained from evaluating the proposed model with 1000 validation instances as follows:

1. Accuracy = 0.959
2. Precision = 0.965
3. Recall = 0.956
4. F1-score = 0.960

From the above metrics, one can conclude that the efficiency of the model is better and in the majority of the cases, we able to identify the accurate sentiment based on movie reviews. From this framework, there are possibilities of about 4.4% for the false negatives and 3.5% for the false positives. If we comprise both the scenarios, there is a possibility of a misclassification rate of 4.1% which is very nominal in the case of sentimental analysis. In this project, we have used the deep learning techniques to train our model. We got some good accuracy in our training results. This project will help us to understand the process of deep learning techniques for twitter sentiment analysis.

CONCLUSION

In conclusion, Twitter sentiment analysis using the BERT model has proven to be a powerful and effective approach for understanding public sentiment and opinions expressed on the social media platform. The utilization of the BERT model, a state-of-the-art transformer-based deep learning architecture, has significantly improved the accuracy and contextual understanding of sentiments present in tweets.

As advancements in deep learning and NLP continue, further research and development in this area hold great potential for enhancing the accuracy and

applicability of Twitter sentiment analysis using the BERT model. With ongoing improvements, sentiment analysis on social media platforms like Twitter will continue to play a pivotal role in shaping the way we understand and interact with public opinions in the digital age.

REFERENCE

- [1] Alsaeedi, Abdullah, and Mohammad Zubair Khan. "A study on sentiment analysis techniques of Twitter data." *International Journal of Advanced Computer Science and Applications* 10.2 (2019): 361-374.
- [2] Chen, Xin, et al. "A novel feature extraction methodology for sentiment analysis of product reviews." *Neural Computing and Applications* 31.10 (2019): 6625-6642.
- [3] Gui, Lin, et al. "Learning representations from the heterogeneous network for sentiment classification of product reviews." *Knowledge-Based Systems* 124 (2017): 34-45.
- [4] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*. IEEE, 2018.
- [5] Ireland, Robert, and Ang Liu. "Application of data analytics for product design: Sentiment analysis of online product reviews." *CIRP Journal of Manufacturing Science and Technology* 23 (2018): 128-144.
- [6] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [7] Kiran, M. Vamsee Krishna, et al. "User-specific product recommendation and rating system by performing sentiment analysis on product reviews." *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2017.
- [8] Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. "Twitter sentiment analysis: The good the bad and the omg!." *Fifth International AAAI Conference on weblogs and social media*. 2011.
- [9] Liang, Ruxia, and Jian-Qiang Wang. "A linguistic intuitionistic cloud decision support model with sentiment analysis for product selection in E-

- commerce." *International Journal of Fuzzy Systems* 21.3 (2019): 963-977.
- [10] M. M. Mostafa, "Clustering halal food consumers: A Twitter sentiment analysis," *Int. J. Mark. Res.*, vol. 61, no. 3, pp. 320–337, 2019.
- [11] Mubarak, Mohamad Syahrul, Adiwijaya, and Muhammad Dwi Aldhi. "Aspect-based sentiment analysis to review products using Naïve Bayes." *AIP Conference Proceedings*. Vol. 1867. No. 1. AIP Publishing LLC, 2017.
- [12] Ramadhani, A. M. and H. S. Goo, "Twitter sentiment analysis using deep learning methods," *Proc. - 2017 7th Int. Annu. Eng. Semin. Ina. 2017*, 2017.
- [13] Raj, Vasundhara. "Sentiment Analysis on Product Reviews." *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. IEEE, 2019.
- [14] Shivaprasad, T. K., and Jyothi Shetty. "Sentiment analysis of product reviews: a review." *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, 2017.
- [15] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," *Expert Syst. Appl.*, vol. 110, pp. 298–310, 2018.
- [16] Yang, Li, et al. "Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning." *IEEE Access* 8 (2020): 23522-23530.
- [17] Zhao Jianqiang and G. Xiaolin, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," vol. 3536, no. c, 2017.
- [18] C. dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proc. 25th Int. Conf. Comput. Linguistics*, Dublin, Ireland, Aug. 2014, pp. 69–78, 2014.
- [19] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, vol. 1. Baltimore, MD, USA, Jun. 2014, pp. 655–666.
- [20] R. Socher et al., "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Seattle, DC, USA, Oct. 2013, pp. 1631–1642.