International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 10 | Oct 2023 www.irjet.net p-ISSN: 2395-0072

# A Intensified Approach On Enhanced Transformer Based Models Using Natural Language Processing

**Vinitha.V[1], Dr.R. Jayanthi[2]**

*[1,2]Department of Computer Science and IT, Jain University, Bangalore, India*

---

**Abstract-** *Sentiment analysis is a common technique in artificial intelligence and natural language processing. Automation of reviews of customer comments on products or services is becoming more popular. Sentiment analysis, which analyses and interprets sentiment expressions utilizing a variety of modalities by capturing the interaction of several modalities, makes understanding human emotions as thorough as possible. As a result it can recognize helpful and contextually appropriate information. This article compares and represents two neural network transformers known as BERT (Bidirectional Encoder Representations from Transformers) and ERNIE (Enhanced Representation through Knowledge Integration). When compared to the BERT model, the ERNIE model has improved representation through knowledge integration, which suggests a potential direction for evaluation and additionally, it analyses the numerous datasets used, emphasizing the challenges and possible applications of this technology.*

***Key Words*: Sentiment analysis, artificial intelligence, natural language processing, representation learning, ERNIE, BERT Transformers.**

## 1. INTRODUCTION

Sentiment analysis is a field that analyses information, evaluations, and opinions using machine learning or natural language processing techniques to ascertain the sentiment behind a statement [1]. The growing use of internet apps has produced a sizable volume of short text data containing the user product and service evaluations as well as comments made by users in social media communications. As a result, user opinions on social media are crucial for several business goals. Big data was developed as a result of social media user reviews were exponential development in quantity and significance [2]. When individuals interact on sharing social media platforms like Twitter a lot of data is frequently created like Facebook where user's share their opinions or ideas [3]. Amazon is a well-known e-commerce platform that enables open product evaluation and reviews from customers. If these reviews were considered to determine whether they are advantageous or not, Customers would have the ability to decide on actions ranging from the purchase of a product such as a laptop, mobile phone, or other electronic devices to the composition of movie reviews and making investments, all of which will have an essential effect on daily life. By performing an automated analysis of consumer input such as comments made in response to surveys and discussions held on social media platforms, businesses may be able to learn what it is that makes customers happy or unhappy. They might expand or modify the kind of services they offer to better serve the requirements of their customers [4]. It is crucial for a range of business applications such as product assessments, political campaigns, product feedback, marketing analysis, and public relations, to understand the opinion of the customer [5]. According to the comments left by customers, past approaches to distinguishing product features and non-aspects using the methodologies that are typically employed were inadequate. Additionally, the study aims to provide an efficient sentiment analysis of consumer product evaluations. A classifier based on the fusion of the BERT model is used to examine the emotional content of customer evaluations. Both models can resolve text polysemy and learn semantic information across contexts. The words are directly converted into vectors with word specificity using the trained language model. The input processing is the enhancement of that of the BERT model. The BERT model is paired with a CNN and an RNN to train on the product review text dataset while the fine-tuning phase is in action. BERT is useful for text classification of product reviews because it produces a stronger classification effect than the conventional method. By examining the BERT model and making the appropriate modifications, the ERNIE model is utilized to solve its drawbacks. The ERNIE model performs better and is more dependable than the BERT model.

## 2. LITERATURE REVIEW

Utilizing sentiment analysis in the text is the primary study on the topic of NLP, understanding the content and perspective of the text benefits greatly from implicit emotion. In social networking sites individuals like posting comments and interacting with others which generates a large number of quick reviews. However, studies have found it difficult to decipher the emotions in summary remarks because of the erratic nature of online rapid comment phrases and the ambiguous word meanings. A new language representation model has been made available by Google AI Language that is built on the Transformers BERT's bidirectional encoder representation [6]. Sentiment analysis is a method [7] for quantitatively detecting and classifying opinions that are communicated in text in particular for determining whether the author has a favorable, negative, or neutral attitude toward a particular topic. Especially those that employ supervised machine learning (ML), several researchers have developed novel techniques for sentiment categorization. Cross-domain sentiment analysis. This study's objective was to assess the effectiveness of the aforementioned machine learning approaches while evaluating and changing feature selection procedures in the Single Domain. In terms of dependability, the most successful machine learning approaches were the MNB, SVM, and SGD algorithms. When it comes to feature selection, TF-IDF is preferred over frequency terms due to the significant reduction it offers (more than 60%) despite the research taking less time. The science of sentiment analysis [8] has grown in importance in the disciplines of During the past ten generations, data mining and natural language processing have been increasingly prominent as a set of methodologies. In recent times, deep neural network (DNN) models have been applied to the task of sentiment analysis, and the results have been positive. ABCDM outperformed six previously suggested neural networks for sentiment analysis on both long and short tweet polarity categorizations. The research used three different Twitter datasets in addition to five different reviews. To extract past and future contexts, two bidirectional LSTM and GRU networks are utilized. from the input text as semantic representation. The gradient issue with the GRU methodology makes it difficult to use this method to predict future context using semantic analysis. The depth of investigation might produce insightful data. We used the Word2vec embedding layer to the phrases to acquire more accurate subjective elements. In the CNN model, increasing the batch size reduces overall losses. In the completely linked hidden layers, a dropout approach was applied, which led to an extra speedup. While decreasing batch size minimizes loss, it also leads to more inaccurate sentiment polarization. Two BERT-based approaches for text categorization have been developed uncased BERT-base and cased BERT-base that have been refined [10] to compare and contrast their respective levels of effectiveness. They gathered information from microblogging sites, particularly Twitter. They carried out tests using two different datasets for sentiment analysis and emotion detection. According to studies, the recommended models can accurately detect emotions with 92% accuracy respectively. They emphasized how well BERT categorizes texts. Use two LSTM models for aspect extraction and sentiment analysis [11]. The effectiveness of the model is assessed using a range of word embeddings. Better outcomes were reached when domain embedding was used as an embedding layer. The accuracy of aspect extraction was 95%, Although social media sentiment analysis is commonly studied using the BERT model, a few researchers have used it to examine the importance and subject of emotion analysis in tourism businesses like hotels and restaurants. Several academics have used the naive Bayes approach to achieve minimal optimization. [12]. The development of the BERT model technology enhances the accuracy of emotion identification. The conventional emotion classification algorithm is unable to differentiate between the psychological meaning of words and due to the sparse wording of comments, polysemy has developed. The BERT model can enhance deep feature capture, increase word recognition accuracy, improve word communication, and enhance classification results. For instance, the ERNIE (enhanced representation via knowledge integration) model, which uses knowledge masking approaches was introduced when it comes to recognizing longer semantics they confirmed that the model can enhance universality and flexibility [14]. The Ernie model was trained with data from a collection of published works and a knowledge map. The early version has excelled in several knowledge-driven tasks, according to experiments and can simultaneously use vocabulary, grammar, and knowledge information [15]. A classifier that utilizes a BERT model fusion and an ERNIE that has been upgraded by the BERT model.

## 3. DISCUSSION

This section gives a brief introduction to well-known transformer-based language representation models Ernie (Enhanced Representation by Knowledge Integration) and BERT (Bidirectional Encoder Representations from Transformers.

The BERT model, also known as the bidirectional encoder representation from the transformers model, is a pretraining language model that is based on deep learning and was initially presented by Google AI. Pretraining language models like neural network language models can directly train a huge number of untagged texts, making them appropriate for a variety of downstream NLP tasks including categorization of text, annotation of text, as well as the resolution of questions automatically. The advantage of using a language model that has been trained is that it only needs to be modified to fulfill the objectives and it does not require additional training. The encoder is the component of the transformer model that represents fundamental building block of the bidirectional language model known as the BERT model. The masked language model pretraining task which uses words as input and context semantic data to predict the masked component serves as a representation of the two-way language model. Google AI developed two BERT models depending on parameter size a model that is based on BERT, as well as a model that is BERT-large for direct human use. The following is the structural diagram of the BERT model.

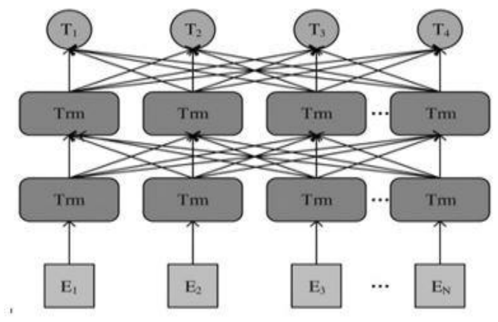


**Fig-1:** Structure of Bert model

The BERT model dissects the text sequence E into token embeddings, segment embeddings, and position embeddings. Furthermore, the text sequence T that the model can comprehend will be produced by vectors that have the same dimensions as one another in each of the three components. Fig-1 depicts the progression of this procedure. The three types of embeddings are known as token embeddings, segment embeddings, and position embeddings. The embedding layers in BERT are detailed below. Making words into fixed-dimension vector representations is the intent of the token embedding layer. A 768-dimensional vector is shown for each word in the BERT model example. Before the input text is transmitted to the token embeddings layer, it is tokenized. Additionally, the phrase that has been tokenized has additional tokens at both the beginning ([CLS]) and the end ([SEP]). These graphical representations serve both as an input representation for classification and as output representations to differentiate between two input texts. Embedding a segment BERT is capable of performing using a pair of input texts, NLP tasks may categorize messages. One example is the categorization of two text snippets that are semantically used to feed, simple concatenation is employed. the introduced mode with the pair of input texts. Positional embeddings word placement in a sentence is described by positional embeddings. The transformer constraint, which prohibits the transformer from recording order information, is controlled by these embeddings. The BERT algorithm can interpret the sentences that are presented as input to positional embeddings. These representations are combined to form a single representation which is added to one another element by element. The Encoder layer of BERT uses this representation as its input representation.

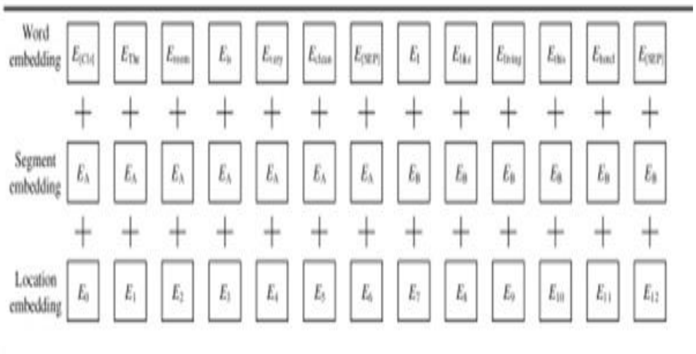


**Fig-2:** BERT input sequence structure diagram in the model

**ERNIE model**

The BERT model structure serves as the foundation for the ERNIE model, which extensively draws on lexical, syntactic, and semantic data. The primary method through which the ERNIE model exceeds the BERT model is the mask mechanism. The lexical, syntactic, and semantic characteristics of the training data aren't fully used to learn modeling, therefore the BERT model guesses up to 15% of the words at random. The modeling object of the BERT model mainly focuses on the original linguistic signal. The language model for the mask in the ERNIE model is a mask mechanism with knowledge. The model can learn the semantic representation and create the link between any two words by modeling semantic data, such as words and phrases. The first masking technique is phrase-based, while the second is entity-based (name, place, organization, product) which has both been enhanced between ERNIE 1.0 to BERT. A phrase or other entity made up of many words is handled as a logical whole in ERNIE and is trained using the entity unit's mask and a word-based mask. One mask can address issues with knowledge dependence and extended semantic dependence during text pre-processing instead of direct knowledge sharing. Since the initial corpus generally contains a sizable quantity of duplicated material that is inappropriate for direct training, it is typically required to arrange it first. The product comment corpus query text contains several irregular objects that include stop words, scrambled codes, and special symbols are all included in the proposed synchronous binary classification model. The pre-processing of words is necessary to minimize the impact of noise and duplicate information and ensure that the text in the emotional corpus has the greatest impact. The product comment message's variable length requires the insertion of a predetermined length or dimension. To satisfy this need the size of the text is configured to truncate and replace. The following have been pre-processed as illustrated in the picture below, the experimental data will be obtained.

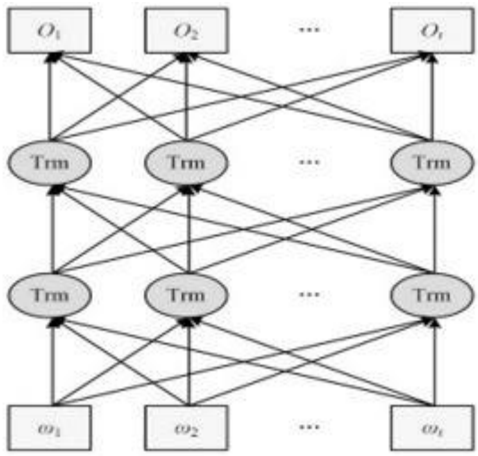


**Fig- 3:** Structure of ERNIE Model

In this study, textual reviews were integrated after data processing using the BERT model since it may be utilized in both traditional and simplified contexts. The BERT model paradigm enables word-level text segmentation. To classify in precise classification, the Whole Word Masking (WWM) approach includes masking each letter in a word. A straightforward vector-based input and output pretraining language model uses the BERT model. This work adds a softmax layer to the output layer for straightforward text categorization of input vectors to achieve text classification. It depicts the precise design by adjusting the values of the three parameters without taking into account the running time and memory leading to a model with a high classification accuracy. The three parameters of the BERT pretraining model's default preset settings have simple binary classification problems including results that fall into the following categories: TP, FP, FN, and TN. TP represents a positive sample, and the forecast outcome is accurate. The following formula is used to determine the related accuracy rate, recall rate, and f1 value.

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * TP}{2 * TP + FP + FN}$$

The model's predictions of sentiment types for the amazon product review texts in the experimental corpus are anticipated to be as accurate as feasible in this study. In model training, the cross entropy between the actual and predicted emotion types serves as the loss function. Model training is done using the algorithm following the loss function that has been established. The loss function in this work has the following precise formula.

$$Loss(p, q) = -\sum_{i=1}^{n} p(x_i)\log(q(x_i))$$

where N, p(xi), and q(xi) stand for the number of emotional categories, actual emotion type probabilities, and expected emotion type probabilities, respectively. To what extent the two distributions approximate one another is determined by cross entropy. The cross entropy is less and the approximation value of the two distributions is larger the closer it is to the real effective type. For training and assessing sentiment analysis algorithms, sentiment analysis uses datasets that are used extensively. A few well-known sentiment analysis datasets are listed below:

**IMDb Movie Reviews:** This set of 50,000 IMDb movie reviews includes both positive and negative ratings. It is a dataset that is frequently used for sentiment analysis applications. For ease of usage, each review has been tokenized and preprocessed.

**Amazon datasets:** It offers with product reviews for a range of items, including books, gadgets, and more. Typically, these reviews are classified as either positive, negative, or neutral.

**Reviews on Yelp:** Yelp provides datasets with reviews about shops, eateries, and other services.

**Twitter Sentiment Analysis:** Sets of tweets with labels indicating whether they are favorable, negative, or neutral. The sentiment of brief text messages can be examined using these datasets.

**Sentiment140 Dataset:** This set of 1.6 million tweets includes labels indicating whether they are favorable or negative. For sentiment analysis of social media data, it is a well-liked dataset.

## 4. CHALLENGES

It faces several problems including biasing, context dependence, slag words, code-mixed data, redundancy, high dimensionality, and domain specificity. These are some of the disadvantages mentioned.

**Cross Domain:** Sentiment categorization is considered as sensitivity of domain since different domains have distinctive ways of expressing opinions. This problem is solved by understanding the characteristics of an unknown domain, sentiment analysis can change according to the circumstances. Words that are used in any one situation might not signify the same thing when used in another.

**High Dimensionality:** It indicates about a significant feature set that lacks from computational problems and invokes for the use of appropriate feature selection techniques.

**Biasing:** Sentiment analysis is commonly used in industries comparable to healthcare, which handles delicate topics like counseling. It's essential to consider bias, particularly when attempting to identify emotional cues in the many customer service calls and marketing leads that come from various racial and socio-economic demographics. Gender, race, age, and other factors can all be sources of bias.

**Context Dependency:** Depending on the subject, several sentimental expressions are used. Even seemingly neutral phrases can carry meaning when combined with other words or sentences, the same phrase can generate negative emotions.

## 5. APPLICATIONS

**Social Media Analysis:** The sentiment of data can be examined via sentiment analysis of social media users towards a particular topic or brand. It can help businesses understand customers' opinions and feedback on their products or services. Here are a few potential uses for sentiment analysis on social media:

**Management of online reputation:** Social media sentiment analysis can be used to track online opinions of a certain product or business. Businesses can learn more about how the public perceives their brand by examining text, photographs, and videos posted on various social media sites. Businesses can use this to pinpoint areas where their reputation needs work and change their messaging accordingly.

**Crisis management:** Sentiment analysis on social media can be used to track public opinion in times of crisis or disaster.

Businesses can learn more about how the public responds to the crisis by examining social media posts and other internet information. This can assist companies in modifying their crisis response plans and messaging, and address any concerns or questions that the public may have.

**Customer service:** In order to serve customers, social media sentiment analysis can be used to examine their comments on various social media sites. Businesses can use this to pinpoint areas for customer service improvement and deal with problems early on.

**Marketing analysis:** Social media sentiment analysis can be used to learn more about how consumers feel about various goods and services. Businesses can learn more about what drives customers to make purchases by studying social media posts, and then modify their marketing efforts as necessary.

**Event Monitoring:** Public opinion towards various events, such as concerts, festivals, and sporting contests, can be done via social media sentiment analysis. Businesses can learn how the public is responding to the event by examining social media posts and they can then modify their strategy accordingly.

Ultimately, social media multimodal sentiment analysis can give businesses insightful information on how the general public feels about their brand or product. Businesses can develop a more thorough grasp of the sentiment towards a given topic and modify their tactics and messaging by examining text across numerous social media platforms.

## 6. CONCLUSION

Despite the enormous growth in smartphone usage, e-commerce has arisen as a new business strategy for companies looking to increase their market share. The sentimental analysis study has been simplified to make it easier for the researchers to choose the best implementation strategies. Since different domains use a variety of languages and cultural contexts, it provides domain adaptability, transfer learning, and multi-task learning. Several associated activities can be completed using these strategies. It has great potential for influencing corporate choices in a variety of industries and for better comprehending consumer attitudes and behaviors. Another approach is the improvement of models for domain adaptation which includes transfer learning and has novel applications in emerging industries like virtual and augmented reality.

## REFERENCES

[1] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, Tang, F., Fu, L., Yao, B., & Xu, W. (2019). Aspect based fine-grained sentiment analysis for online reviews. *Inf. Sci., 488*, 190-204.

[2] Rezaeinia, S. M., Rahmani, R., Ghodsi, A., & Veisi, H. (2019). Sentiment analysis based on improved pre-trained word embeddings. *Expert Systems with Applications*, *117*, 139-147.

[3] Behera, R. K., Jena, M., Rath, S. K., & Misra, S. (2021). Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. *Information Processing & Management*, *58*(1), 102435.

[4] Shaheen, M., Awan, S. M., Hussain, N., & Gondal, Z. A. (2019). Sentiment analysis on mobile phone reviews using supervised learning techniques. *International Journal of Modern Education and Computer Science*, *11*(7), 32.

[5] Aziz, A. A., & Starkey, A. (2019). Predicting supervise machine learning performances for sentiment analysis using contextual-based approaches. *IEEE Access*, *8*, 17722-17733.

[6] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[7] Aziz, A. A., Starkey, A., & Bannerman, M. C. (2017, September). Evaluating cross domain sentiment analysis using supervised machine learning techniques. In *2017 Intelligent Systems Conference (IntelliSys)* (pp. 689-696). IEEE.

[8] Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Future Generation Computer Systems*, *115*, 279-294.

[9] Raviya, K., & Vennila, S. M. (2021). Deep cnn with svm-hybrid model for sentence-based document level sentiment analysis using subjectivity detection. *ICTACT Journal On Soft Computing*, *11*(3).

[10] Chiorrini, A., Diamantini, C., Mircoli, A., & Potena, D. (2021, March). Emotion and sentiment analysis of tweets using BERT. In *EDBT/ICDT Workshops* (Vol. 3).

[11] Sindhu, I., Daudpota, S. M., Badar, K., Bakhtyar, M., Baber, J., & Nurunnabi, M. (2019). Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. *IEEE Access*, *7*, 108729-108741.

[12] Zvarevashe, K., & Olugbara, O. O. (2018, March). A framework for sentiment analysis with opinion mining of hotel reviews. In *2018 Conference on information communications technology and society (ICTAS)* (pp. 1-4). IEEE.

[13] Lu, Q., Zhu, Z., Xu, F., Zhang, D., Wu, W., & Guo, Q. (2020). Bi-gru sentiment classification for chinese based on grammar rules and bert. *International Journal of Computational Intelligence Systems*, *13*(1), 538-548.

[14] Sun, Y., Wang, S., Li, Y., Feng, S., Chen, X., Zhang, H., ... & Wu, H. (2019). Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

[15] Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., & Liu, Q. (2019). ERNIE: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.

[16] Meng, W., Wei, Y., Liu, P., Zhu, Z., & Yin, H. (2019). Aspect based sentiment analysis with feature enhanced attention CNN-BiLSTM. *IEEE Access*, *7*, 167240-167249.

[17] Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-dependent sentiment classification with BERT. *Ieee Access*, *7*, 154290-154299.

[18] Kokab, S. T., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, *14*, 100157.

[19] Xu, N. (2017, July). Analyzing multimodal public sentiment based on hierarchical semantic attentional network. In *2017 IEEE international conference on intelligence and security informatics (ISI)* (pp. 152-154).

[20] Yuan, J., Mcdonough, S., You, Q., & Luo, J. (2013, August). Sentribute: image sentiment analysis from a mid-level perspective. In *Proceedings of the second international workshop on issues of sentiment discovery and opinion mining* (pp. 1-8)