

## App store bugs-review classification using BERT- DNN model

Mrs.Sabitha .P<sup>a</sup>, Mr.Sharan Nixon<sup>b</sup>, Mr.Debin Jose<sup>c</sup>, Mr. Abhijith KP <sup>d</sup>

<sup>a</sup>Asst. Professor in CSE Department, SRM University, Chennai, Tamil Nadu, India

<sup>b,c,d</sup> UG Scholars, CSE Department, SRM University, Chennai, Tamil Nadu, India

**Article History:** Received: 5 April 2021; Accepted: 14 May 2021; Published online: 22 June 2021

**Abstract:** There so many applications available on the internet (Especially in Google play store), which many are using for all types of different purposes. Many users gives their reviews about the applications they used like their experience, the issues they face, the updates which will be helpful for them, etc. All of these reviews should be analysed so that the developer can improve the applications by adding required updates needed by the users, and fixing issues faced by the users. For this purpose we need to analyse the reviews in an efficient manner which will save more time and improve their applications. In this project we are doing classification of bugs in reviews with the help of encoder based BERT(Bidirectional Encoder Representations from Transformers) model,which have been very popular in natural language processing, has been widely used now a days. This will help us to classify the reviews in a manner that we can identify positive and negative reviews which will help the developers to find the bugs. The reports that are under negative and neutral reviews are considered for bugs. It is found that BERT model is much efficient in text classification so it will be very well suitable for review based classification. There four different phases in this project first one is the data exploration, then we pre-process the data, and then do feature selection, finally we classify the reviews using BERT model.

**Keywords:** Transformers, Self attention mechanism, encoder, Bert, Natural language processing, bugs classification.

### 1. Introduction

In our daily life internet plays a vital role. Now the technology is advanced,we get all the information in our fingertip. We can implement all the works with the help of internet. These days, so many applications and other platforms are available in online store, which are different on their quality and performance. One of thing is that their reviews are readily available. App reviews in online platforms can be useful for new buyers who select high-quality products as well as for the sellers to enhance the app qualities and the efficiency. Sentimental analysis for bug reviews is mainly done with the help of machine learning and deep learning. The reviews may include positive, negative and also it contains some neutral reviews. Now the quality and accuracy of sentimental analysis is improved using so many methods,one of the most efficient method to improve the accuracy of sentimental analysis is BERT model. The model contains a layered representation input to encode the text of reviews that given by the users in different online stores. Texts are encoded and extracted.After this process next is DNN semantic extraction that was utilized to extract the local features and common features of these reviews that contain vectors of texts. Here BERT semantic extraction layer to extract and take the main important and remarkable reviews of text vectors. Deep Neural Networks (DNN) has a feature of automatically extracting the text from user review and it gives an optimal solution that contains the artificial form of solution of classification of reviews. The negative review and neutral reviews are taken for bugs' consideration. This help the developers to work efficiently and to help save their time.

### 2. Relatedworks

So many approaches of sentimental analysis are already done on the basic of BERT model. That gives well defined solutions. The paper "Aspect extraction from tourist spot review based on BERT"[2],in this paper they proposed a model that analysis the review of users which they entered in many tourist related website. The development and demand of tourism is improved day by day,so many websites are available for tourists to search about that place where they plan to go. Bidirectional encoder representation from transformers model is used to extract and encodes these reviews in the form of text vectors, in this project they added a special feature of Indonesian language for the further enhancement and study. In this model BERT improve the accuracy and efficiency.The proposal of this project is mainly divided in to two parts, one main process is addition of the pre-processing phase,it is very challenging and further stage is pre-training process,that means it is BERT Multilanguage base model with domain specific database.The usage of multilingual base model with Indonesian language as one of main dataset language .The Indonesian particular model for BERT is still exist and it also enhance the accuracy of model .Which leads to the limitation of computing power and the availability of Indonesian language database.

"A Commodity Review Sentiment Analysis Based on BERT-CNN Model"[3], in this paper they proposed a model which takes the commodity reviews form users which they given in e-commerce websites,in this review contains both positive,negative and also neutral reviews.In this model they try to improve the quality and efficiency of e-commerce apps and then it impact the demand of these kinds of apps in market.Now quality and

accuracy of sentimental analysis is improved using so many methods, one of the most efficient method to improve the accuracy of sentimental analysis is BERT model. In this model it contains a layered representation input in to the model that encode the text of reviews that given by the users who use mainly e-commerce apps. Then these texts are encoded and extracted. After this process next is CNN semantic extraction, that was utilized to extract the local features and common features of these reviews that contains vectors of texts. So in this paper sentimental analysis based on users review done in BERT model. This method work on the principle of deep learning and also it has a feature of automatically extract the text from user review and it gives an optimal solution.

“The Automatic Text Classification Method Based on BERT and Feature Union”[3], in this paper they proposing classification of text automatically on basis of BERT (Bidirectional Encoder Representation from Transformers), which is used as a pre-training model. And they are also adding feature union to this process which will give more accuracy to this model. Classification of text is actually a task meant for NLP (Natural Language processing), which is why they are using BERT model which is actually transform architecture. Here they are trying to make full use of CNN by transforming character level embedding with the help of BERT model. Here they are introducing new idea for text classification which will be very much useful for extracting relevant data from a text. Its usage is actually more and more for gathering information’s from a text.

### 3. Proposed System

The proposed system uses advanced machine learning algorithm known as BERT (Bi-directional encoder representation from transformers). This machine learning algorithm is best suited for the Natural Language Processing. Since this project uses app store reviews dataset to classify reviews that contains bugs or not, BERT algorithm uses shallow deep decoder to generate different word sequence. It can also generate any missing sequence. Compared to LSTM, Bert model takes less training time. Since Bert can learn bidirectional data other than traditional left to right or right to left methods, it has more advantages compared to other algorithms.

In this system, supervised learning technique is used for training. First the system will train to classify a whether it is a neutral, negative or positive review. All the reviews that classifies as negative or neutral are taken into consideration for reviews with bugs. The loss function for this model used was spares categorical cross entropy, since the output was classification. The final layers are of the model are connected to a fully connected deep neural network. The output size of the neural network has 3 nodes for which each node represents negative positive and neutral. Soft-max function is used at the end to take the probabilistic value of each node and the node with max probability is taken as the output from the model.

### 4. Implementation

#### A. Data Exploration:

The dataset contains 64,296 reviews of different apps and its user reviews. The dataset had many Nan values and therefore most of the rows containing Nan Values were dropped from the data frame. After dropping the Nan values, the dataset contains 37,427 rows in which 23,998 are positive reviews, 8,271 negative and 5,158 neutral reviews.

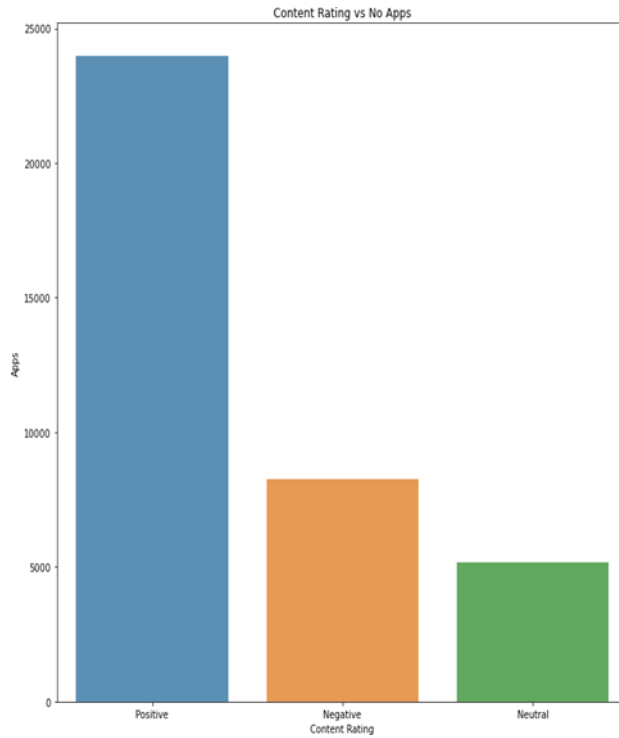


Fig-1

In order to further explore the dataset, the word cloud is plotted. The word cloud plots all unique words in the reviews. Also it the size of the word in the word cloud is based on the number of times it is repeated in the dataset.

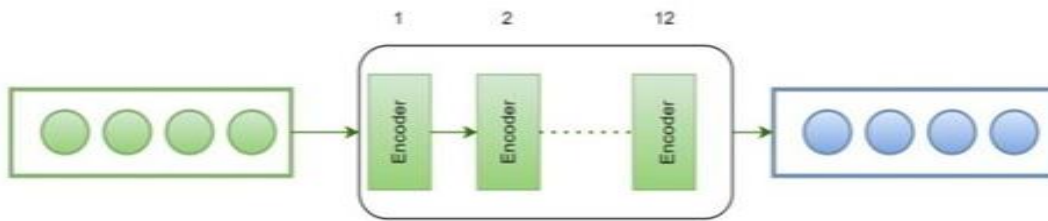


Fig-2

The Fig-2 contains a word-cloud of all positive reviews. Word cloud shows the important keyword in all positive reviews.



**B. Model architecture:**

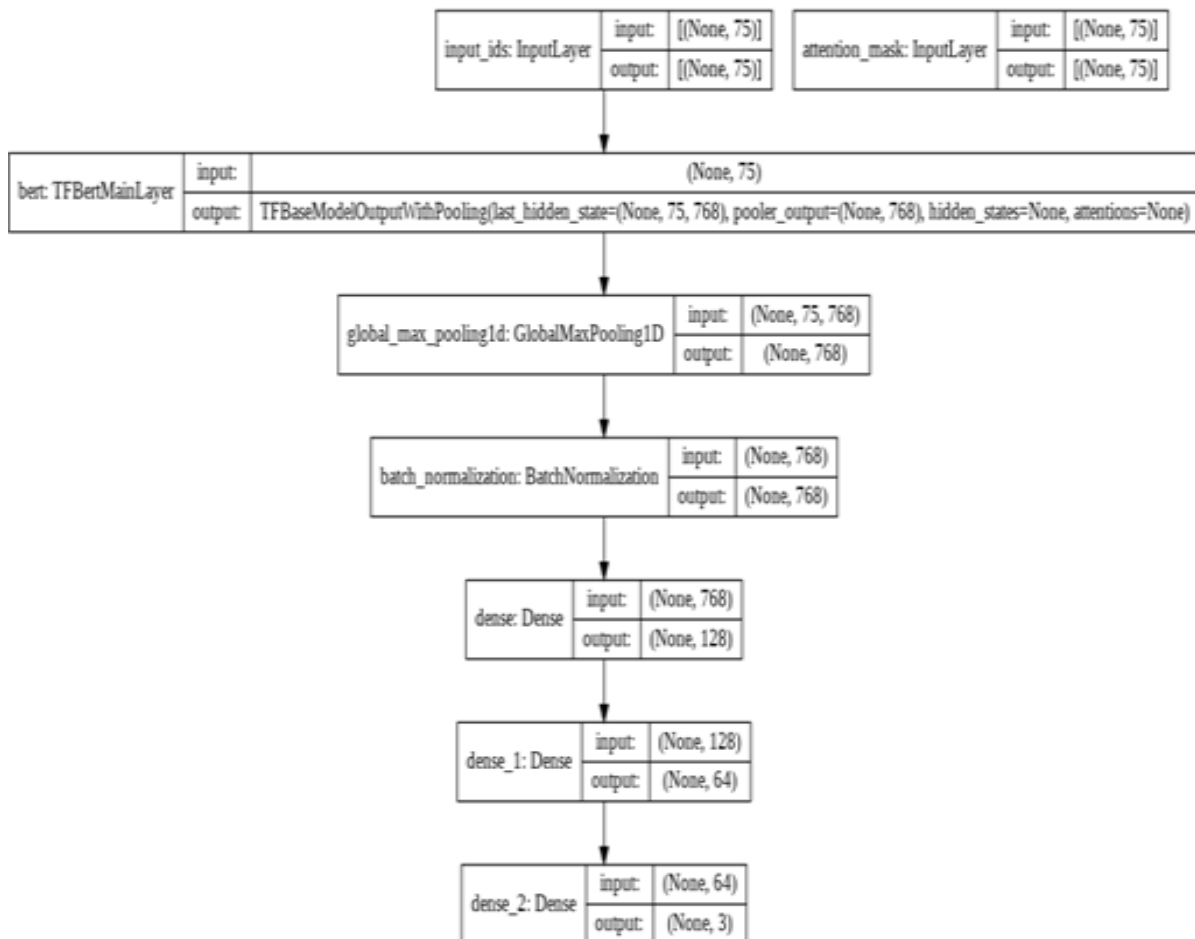


**Fig-5**

The Bert model used in this model contains only encoders. Encoder encodes the entire input sequence to fixed length and also generates embedding's for each word. The embedding is the encapsulation of meaning of that word. Therefore similar words will have closer embedding values. Bert input embedding's consist of three things, token embedding's, segment embedding's, position embedding's.

Bert masks intermediate token in sequence for prediction task for bi directional encoding majority masking strategy in Bert replaces keywords with masks tokens. In this model, the self-attention mechanism allows the model to understand the necessary keywords and their positions in the sentence. The encoder learns about the language structure and its context.

As shown in the fig-6, the output node of the Bert model consists of 768 nodes. Therefore a dense neural network containing 768 nodes is used to take the output values from the Bert model. This layer is connected to the next dense\_1 layer containing 128 nodes and the output of this node is connected to dense\_2 layer containing 64 nodes. The output layer consists of 3 nodes since the output of the prediction is Positive, Negative and Neural. Soft-max function is used in the output layer to convert the outputs to probability values. The model consists of 108,420,227 total parameters in which trainable parameters is 108,419 and non-trainable parameter is 108,311,808.



**Fig-6**

**C. Pre-processing:**

The model two things as input for each reviews. An input indices and attention mask. The input indices contain sequence tokens in the vocabulary while the attention mask contains values 0 and 1 to avoid padding token indices. Therefore the reviews data needs to be converted. It is done using the Bert tokenizer. The labelled data i.e. positive, negative and neutral is converted into 1, 0 and 2 respectively. About 90 percent of the data was taken as the training data and 10 percent data was taken for testing.

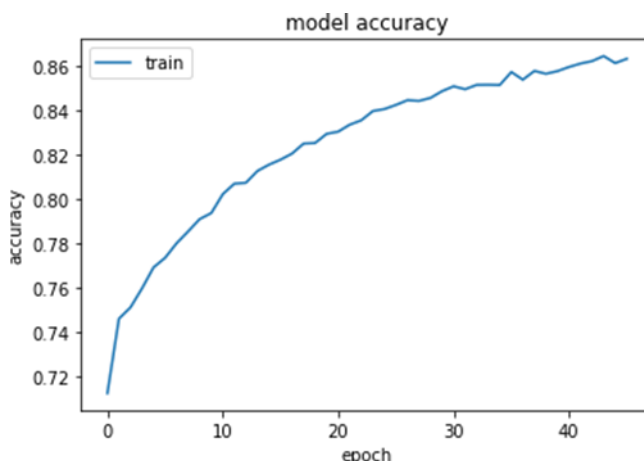
**D. Training the model:**

The model is trained as a classification model. Therefore loss function used is sparse categorical cross entropy. The optimizer uses for training was Adam. In order to fully train the model, early stopping function was used to train the model. In early stopping, the parameters used were, min-delta and patience was set to

2. Mode was not changed from ‘auto’ and restore best weights were set to ‘True’. The number of epochs to train was set to 55. Each epoch consisted of 1170 training steps since the data was shuffled and made into 32 batches. The shuffle parameter was set to 100,000. Testing dataset was used for accuracy validation data.

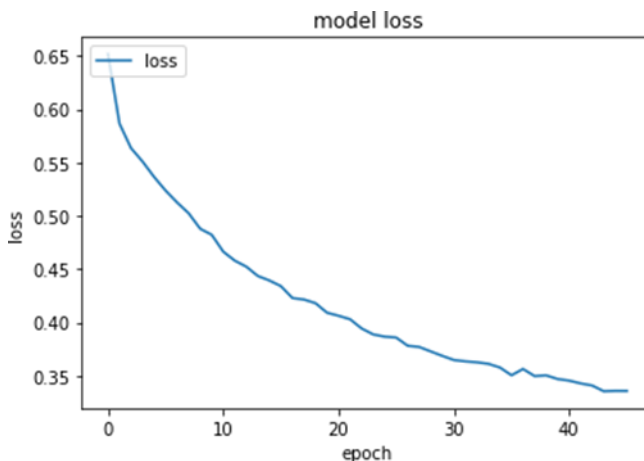
**5. Results**

The model stopped training at epoch number 46 since the early stopping function was used. It took about 3 hours to train the model using GPU from Google co- lab. The model training history is shown in fig-7.



**Fig-7**

The training loss for the model is shown in the fig-8. The final loss for the model was 0.33.



**Fig-8**

**6. Conclusion**

This paper presents a different approach in predicting bugs from apps using an advanced natural language processing algorithm like Bert. Even though 86 percent accuracy was attained, the variance and bias in the dataset

is still in question. After plotting the word-cloud of dataset, there seems to be common words in both negative and positive. With a rich quality of dataset, the accuracy can go above 90 percent.

## 7. Future Scope

Natural language processing is relatively new field and is still in its initial development stages from fully developed language processing system. The applications of natural language processing are very vast, opening different door into the future. There are still many developments in this field to occur. It is possible to make a system which can fully recognize different types of bugs from reviews like Humans.

## References

1. Wenting Li, Shangbing Gao, "The Automatic Text Classification Method Based on BERT and Feature Union", 4-6 Dec. 2019, DOI: 10.1109/ICPADS47876.2019.00114.
2. Muhamad Rizky Yanuar, Shun Shiramatsu, "Aspect Extraction for Tourist Spot Review in Indonesian Language using BERT", 16 April 2020, DOI: 10.1109/ICAIC48513.2020.9065263.
3. Junchao Dong, Feijuan He, Yunchuan Guo, Huibing Zhang, "A Commodity Review Sentiment Analysis Based on BERT-CNN Model", 16 June 2020, DOI: 10.1109/ICCCS49078.2020.9118434.
4. C. S. Khoo and S. B. Johnkhan, "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons," *Jour. Inform. Scien.*, vol. 44, no. 4, pp. 491-511, Aug. 2018, DOI:10.1177/0165551517703514.
5. R. L. Rosa, D. Z. Rodriguez, and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks," *IEEE Trans. Consum. Electron.*, vol. 61, no. 3, pp. 359-367, Aug. 2015, DOI:10.1109/TCE.2015.7298296
6. M. Colhon, C. Bădică, and A. Şendre, "Relating the opinion holder and the review accuracy in sentiment analysis of tourist reviews," in *Int. Conf. Knowledge Sci., Eng. and Manage.*, 2014, pp. 246-257, DOI:10.1007/978-3-319-12096-6\_22
7. M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in *Proc. 9th Int. Workshop Semantic Evaluation*, 2015, pp. 486-495.
8. L. Zhang and B. Liu, "Aspect and entity extraction for opinion mining," in *Data Mining Knowledge Discovery For Big Data*, Berlin, Germany: Springer, 2014, pp. 1-40, DOI: 10.1007/978-3-642-40837-3\_1
9. S. Schuster and C. D. Manning, "Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks," in *LREC*, May. 2016, pp. 23-28
10. New Google Play Store greatly simplifies permissions. <http://www.androidcentral.com/new-google-play-store-4820-greatly-simplifies-permissions>, 2014.