Transformer based Contextual Model for Sentiment Analysis of Customer Reviews: A Fine-tuned BERT

A Sequence Learning BERT Model for Sentiment Analysis

Ashok Kumar Durairaj[1], Anandan Chinnalagu[2]

Department of Computer Science, Govt. Arts College
(Affiliated to Bharathidasan University Tiruchirappalli)
Kulithali, Karur, India

Abstract—The Bidirectional Encoder Representations from Transformers (BERT) is a state-of-the-art language model used for multiple natural language processing tasks and sequential modeling applications. The accuracy of predictions from context-based sentiment and analysis of customer review data from various social media platforms are challenging and time-consuming tasks due to the high volumes of unstructured data. In recent years, more research has been conducted based on the recurrent neural network algorithm, Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM) as well as hybrid, neutral, and traditional text classification algorithms. This paper presents our experimental research work to overcome these known challenges of the sentiment analysis models, such as its performance, accuracy, and context-based predictions. We've proposed a fine-tuned BERT model to predict customer sentiments through the utilization of customer reviews from Twitter, IMDB Movie Reviews, Yelp, Amazon. In addition, we compared the results of the proposed model with our custom Linear Support Vector Machine (LSVM), fastText, BiLSTM and hybrid fastText-BiLSTM models, as well as presented a comparative analysis dashboard report. This experiment result shows that the proposed model performs better than other models with respect to various performance measures.

Keywords—*Transformers model; BERT; sequential model; deep learning; RNN; LSVM; LSTM; BiLSTM; fastText*

I. INTRODUCTION

The RNN, LSTM, gated recurrent neural network (GRNN) and BiLSTM models are some of the various sequence models [1] for NLP tasks, language modeling, and machine translation. The Google research team introduced (Year 2017) the first Transformer and the first transduction model that replaces the recurrent layer with attention. The first Transformer model was used for language translation (English to France and English to German) tasks and the results showed that it outperformed all other neural model architecture with convolutional or recurrent layers [2]. Fig. 1 shows the Transformer architecture and the build blocks of BERT model. In this Transformer architecture, input sequence $(x_1, x_2, x_3, \ldots, x_n)$ mapped for symbol representation of encoder to the continuous sequence representation of decoder $(z_1, z_2, z_3, \ldots, z_n) = z$ the z decoder generates the $(y_1, y_2, y_3, \ldots, y_m)$ sequence symbol of an element output at a time.

Model generates the symbols based on the output of previously generated symbols as an additional input, it uses the auto-regressive method at each step for next generation of symbols [2]. The overall Transformer architecture is based on self-attention with fully connected encoder and decode.

BERT is a pre-trained, an open-sourced (Year 2018) [3] and transformer-based language model from Google. It's designed to pre-train bidirectional (left and right) text representations from unlabeled text [4]. The $BERT_{BASE}$ and $BERT_{LARGE}$ are two original models. The base model consists of 12 Encoders and bidirectional self-attention, while the large model consists of 24 Encoders and 16 bidirectional heads. BERT model is pre-trained on 800 million words from BooksCorpus and English Wikipedia's unlabeled text of 2.5 billion words. As BERT is pre-trained with large unlabeled text datasets, this model can be easily fine-tuned for small datasets that are specific to an NLP task like sentiment prediction on customer or employees' reviews and question-answer system for chatbot applications.

Fig. 2 shows the neural network architecture of BERT's deep bidirectional and OpenAI GPT's unidirectional (Left-to-Right) contextual models [1], in which the unidirectional model generates a representation for each word based on other words in the same sentence. The BERT bidirectional model represents both the previous and next context in a sentence. However, the context free Word2vec and Glove models generate a word representation based on each word in the vocabulary.

There are many organizations that rely on reviews to improve customer experience and increase revenue of their products and services. Positive sentiments are one of the key factors for the success of several online businesses. However, determining the context of the review, polarity, and sentiment in textual content of customer reviews remain a challenge. The custom and hybrid deep learning models (LSTM, BiLSTM, fastText and fastText-BiLSTM) perform higher in textual datasets compare to traditional models [Naive Bays (NB), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Support Vector Machines (SVM)].
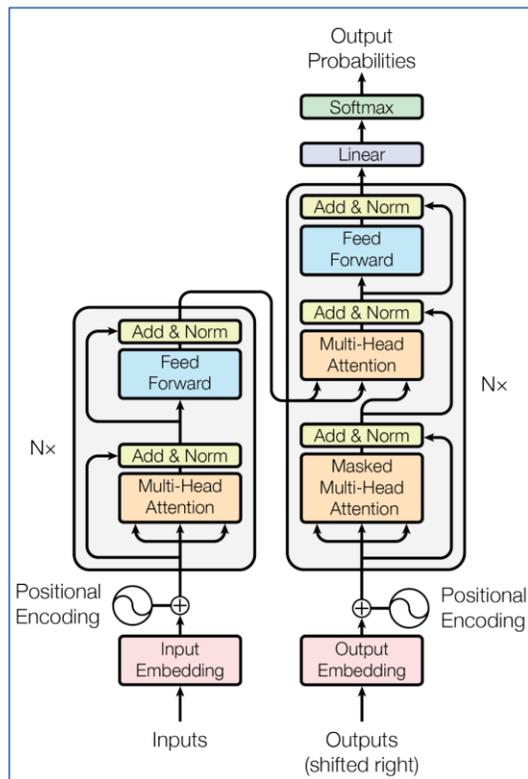
Fig. 1. Multi-layer Transformer Architecture.
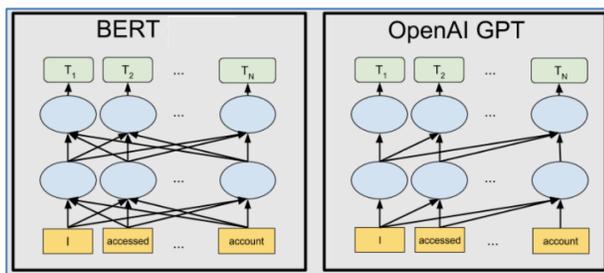
Fig. 2. BERT and OpenAI GPT Neural Network Architecture.

Evaluating Transformer based models and comparing their results with other state-of-the-art LTSM models are important for identifying the best model to utilize for Sentiment Analysis applications. Several researchers have used traditional models for sentiment analysis on datasets of customer reviews and discussed about their experience and issues occurred from it. These issues include time consumption of tasks such as data pre-processing, preparation for testing and training the datasets, as well as problems in performance and accuracy. We identified the gap in leveraging the pre-trained BERT model for datasets of customer reviews and found that the pre-trained model solves the problems and issues that traditional models have.

The pre-trained BERT model's performance, accuracy and approach motivated the authors to experiment with this model for customer sentiment analysis. The main objectives and contributions are as follows;

- Overcome known challenges of SA model performance, accuracy and context-based prediction.

- Train BERT-base-cased model on Twitter, IMDB Movie Reviews, Yelp, and Amazon customer reviews datasets to improve the accuracy and performance of the model.

- Evaluate the custom deep learning sequential model of BiLSTM, hybrid fastText-BiLSTM model and linear models of LSVM, fastText models using the same datasets.

- Compare the results of the BERT model with the results of the deep learning sequential and linear model.

- Customize the data pre-processing steps for hybrid and linear model training.

- Fine-tune the hyperparameters for fastText-BiLSTM models.

This paper presented with several recent BERT and SA related research papers reviews and major contributions from various researchers in Section II. The literature reviews of BERT, fastText, BiLSTM, and LSVM models are presented in Section III. The experimental setting and model evaluation results are discussed in the Section IV. In Section V, concluded this paper with model results, findings of this research work and future work.

II. RELATED WORK

Recently, many researchers evaluated BERT model for many NLP tasks. In this section, presented most recent research papers on SA and pre-trained BERT models.

In 2021, [5] evaluated RNN+LSTM and BERT model for sentiment analysis to detect cyberbullying based on Twitters Spanish language dataset. The evaluation results show the accuracy and performance of the BERT model outperformed RNN+LSTM by 20%. Based on evaluation the bert-base-multilingual-uncased and Bert-large-uncased models show more accuracy. However, BERT model requires higher configuration of the computational environment. A challenge researchers face is finding the investment on a grand-scale infrastructure to train the models. There are many BERT-based SA experimental research, case studies and review papers presented for Arabic aspect-based [6], Italian Twitter SA [7] and Bangla-English Machine Translation [8].

In 2021, [9] present a large-scale open source pre-trained BERTweet model for English Tweets. BERTweet used for Part of speech (POS), recognition of Named entity and text classifications. Experimental result shows that it outperforms XLM-Rbase and RoBERTabse models, all these models are having a same architecture of BERT-base. There are several models available as open-sourced, whereas other models are inaccessible to be customized for commercial use.

In 2020, [10] presented a research article about a question answering (QA) system based on LSTM, multilingual BERT and BERT+vnKG models. They had used crafted Vietnam tourism QA dataset for this experiment and evaluated three models with the same data. The experimental result shows the proposed model outperformed other model in terms of time and the accuracy.

In 2021, [11] conducted an Aspect-based SA study on consumers product reviews data. They have proposed two BERT models for aspect extraction, sentiment classification using parallel and hierarchical aggregation methods based on hierarchical transformer model [12]. The following Fig. 3 and Fig. 4 show the parallel and hierarchical models. Study result shows that applying their proposed model approach improved the model performance and eliminate the need of further model training.
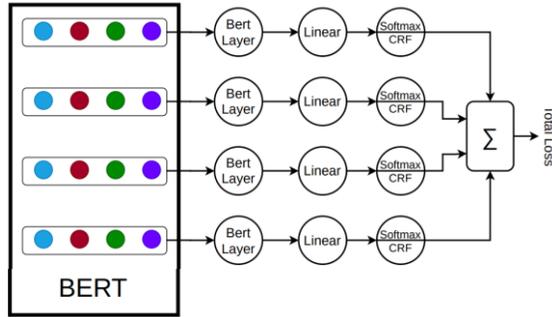
Fig. 3. Parallel Agrregation Model.
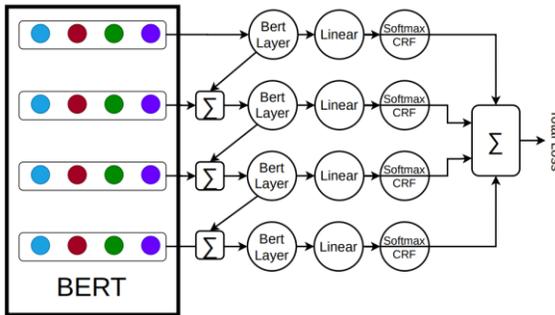
Fig. 4. Hierarchical Aggregation Model.

In 2021, [13] presented a research paper on attentions and hidden representation of learned pre-trained BERT models for aspect-based SA (ABSA) on reviews datasets. This research study found that BERT uses the aspect representation dedicated to semantics of the domain and self-attention head encode the context of the words. In 2021, [14] conduct a experiment based on Aspect-level SA (ABSA) using ALM-BERT model on consumer datasets, ALM-BERT model show the better performance than the other models.

In 2021, [15] presented BERT-based SA models research paper, in this research analyzed Github, Jira web portal comments and Stack Overflow posting datasets related to Software Engineering. comments. Authors used vanilla BERT, fine-tuned ensemble BERT and compressed DistilBERT models [16] and based on F1 score measure the ensemble and compressed BERT models improved performance by 6-12% over other model. Another research study conduted by [17] on GitHub dataset and the exprimental results show 94% accuracy of emotion classification.

In 2021, [18] published an article of SA models comparative analysis od machine and deep learnings models, SA research work carried out by using Naive Bayes (NB), Support Vector Machine (SVM), LSTM and BERT-based

uncased model on consumer reviews and rating dataset from Amazon, Flipkart e-commerce platforms. Binary classification of consumer sentiment is categorized as positive and negative. This research study results show that deep learning BERT uncased model improved performance, and higher accuracy of sentiment predictions compared [19][20] to other machine learning models.

III. PROPOSED MODEL AND METHODOLOGY

Proposed sentiment analysis frameworks and models' architectures present with data pre-processing steps in the following section.

A. Data Pre-processing

We've developed python scripts for data pre-processing steps, customized the scripts and steps sequences based on dataset. Our custom script consists of data cleansing, reduction of unwanted data, data transformation and validations steps.
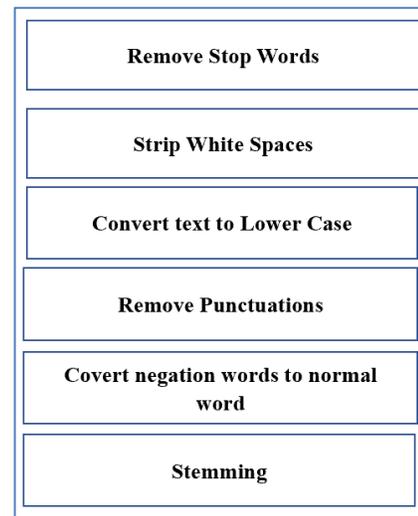
Fig. 5. Data Pre-processing Steps.

The above Fig. 5 shows data pre-processing steps. For this experiment, Twitter, IMDB Movie Reviews, Yelp, Amazon customer reviews datasets are used for training and testing the models. To improve the data quality, model performance and accuracy, performed the following pre-processing tasks: converted text to lowercasing text and removed stop words, converted negation words to normal word, removed white spaces, special characters, punctuations, stripping recurring headers from the text and stemming.

B. Proposed BERT Model

We propose fine-tuned BERT-base-cased model for sentiment prediction on customer reviews datasets. The advantage of BERT model is that it was pre-trained on large corpus of raw texts data. These models are used for Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) and these pre-trained models can be fine-tuned for downstream applications [2]. BERT-base fine-tuned SA model framework is represented in Fig. 6. This framework consists of the following four components, input data pre-processing, Tokenization of input text for model, BERT-base-cased model and classification layer.
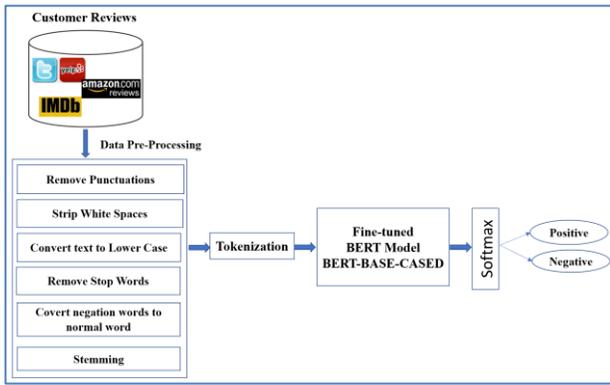
Fig. 6.   BERT-base-cased Sentiment Analysis Framework.

The input layer generated tokens for given pre-processed text, the following Fig. 7 shows the embedding layers of tokanization of sentences. The Token embeddings is the sum of segement and position embeddings. These embeedings are learned specific token from WordPiece token vocabulary [4].

To train the model on customer review dataset, we follow the pre-train procedure with Adam optimzer from Hugging Face. following Fig. 8 shows the training history of BERTmodel on our datset, the model training result show the close to 100% accuracy after 10 epochs. We've evaluated the pre-trained model and the evalution result show 95% accuracy on our test dataset.

Fine-tuned the hyperparameters for higher accuracy and performance for customer reviews datasets. The Softmax function applied in the output layer to get the predicted probabilities of sentiment values. Here are the questions for probabilities and Softmax.

$$Probability = \frac{Numerator}{Denominator} \tag{1}$$
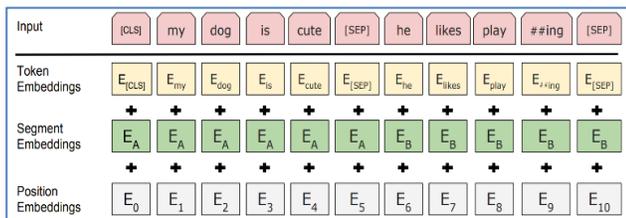
Fig. 7.   Input Representation of BERT Model.
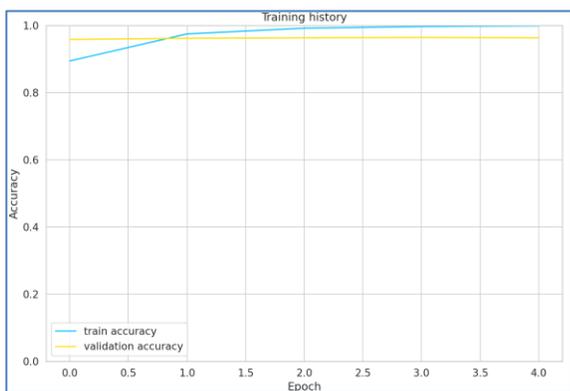
Fig. 8.   Model Training History.

Softmax performs the transform on $x_1, x_2, x_3 \dots x_n$ numbers.

$$softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{2}$$

C. FastText, BiLSTM and FastText-BiLSTM Models

The Fig. 9 shows the neural networks SA framework used for this experimental research. These models are based on fastText and LSTM architectures.

FastText introduce by Facebook AI research (FAIR) lab for text representation and classification, it's a lightweight method and works on generic hardware with multicore CPU, the fastText models show faster performance, accuracy during model training and evaluation [21]. A new extension of the continuous skipgram and Continuous Bag of Words (CBOW) model like word2vec word embedding is introduced by fastText, where each word is represented as a bag of character n-grams. fastText model is Pre-Trained on Wikipedia dataset (294 languages) [22]. To compare with BERT model result, we evaluate customer review dataset using fastText model.

Authors build a two novel multilayer sequence processing custom models using BiLSTM and hybrid fastText-BiLSTM. it consists of two LSTM units and multilayers, one unit taking the input in a forward direction and other unit taking the input in a backward direction [23]. In this experiment, our main goal is to compare the BERT model result with authors The Hybrid fastText-BiLSTM and custom multilayer BiLSTM. Models consist of 196 memory units, 128 Embedding Layer, 5 Dense Layer and Softmax activation function at output layer. The LSTM unit consists of input, output and forget gates and these three gates are the activation of sigmoid function; the sigmoid output value is between 0 and 1, when gates are blocked the value is 0 and gates allow the input to pass through when the value is 1. The following are the equations of sigmoid and input, output and forget gates.

$$sig(t) = \frac{1}{1+e^{-t}} \tag{3}$$

$$i_t = \sigma(\omega_i[h_{t-1}, x_t] + b_i) \tag{4}$$

$$o_t = \sigma(\omega_f[h_{t-1}, x_t] + b_o) \tag{5}$$

$$f_t = \sigma(\omega_o[h_{t-1}, x_t] + b_f) \tag{6}$$

σ represents sigmoid function,

x_t represents input at current timestamp.

h_{t-1} represents LSTM block output of previous state at timestamp t-1.

$\omega_i, \omega_f, and \ \omega_o$ are representing weight of input, forget and output gates.

b_i, b_o and b_f are representing bias for input, output, forget gates.
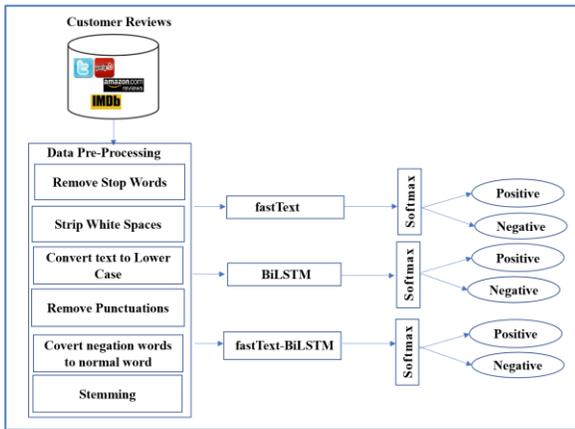
Fig. 9.   FastText, BiLSTM and FastText-BiLSTM Models.

D. LSVM Model

Linear SVM (LSVM) is a most frequently used supervised learning algorithm for classification problems. SVM has two types of classifiers Linear [24] and Non-Linear. In Linear Classifier separates the data in a liner order and a data point considered as a p-dimensional vector, but the best hyperplane consider one which maximize the margin. The text can be categorized linearly using the SVM's linear kernel and LSVM [25] works well with lot of features and also less parameters to optimize for training the model. We used linear kernel for this experiment. Fig. 10 shows the LSVM architecture and SA framework used for this experiment.
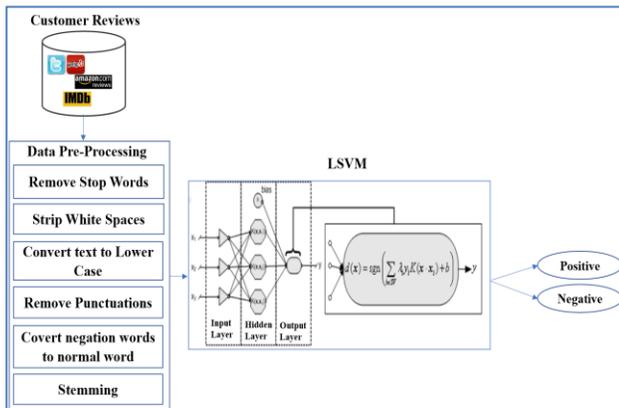
Fig. 10. Linear SVM (LSVM) Model.

Here is the equation for linear kernel function $k(x, y)$, given by the inner product $< x, y >$, and c represent the optional constant.

$$k(x, y) = x^T + c \qquad (7)$$

IV. Experimental Setting and Results

There are four models and two environments used for this experiment. The data source, dataset, environment settings, fine-tuned hyperparameters and experimental results are presented in the following sub-sections.

A. Data Source and Dataset

There are multiple data sources are available online for customer reviews data, we have obtained publicly available

Twitter, IMDB Movie Reviews, Yelp and Amazon customer reviews datasets from Kaggle.com. Total 778631 customer reviews finalized after preprocessing the raw dataset and split the dataset 545041 (70%) for training and 233590 (30%) for testing the models. The data pre-processing steps are explained in the proposed model Section III. Models have trained and tested after fine-tuning the hyper parameters.

B. Experimental Environments

Training and evaluation of the BERT model is used through the Google Colaboratory cloud environment, while a standard server is used for training and evaluating the fastText, LSVM and SA-BLSTM models. Table I lists the details and components used for this experiment.

TABLE I.   Experimental Environments

Environment #1	Description
Model	BERT-base-cased
Transformer	Hugging face
Server Configuration	Google Colab Cloud
Programming Language & Tool	PyTorch, Python 3.8.8 Jupyter Notebook 6.3.0
Environment #2	**Description**
Model	fastText, LSVM and SA-BLSTM
Server Configuration	Windows 64-bit Operating System with Intel core i7 processor, 16 GB Memory.
Word Embedding	fastText Library and Keras Encoder
Programming Language & Tool	Python 3.8.8 Jupyter Notebook 6.3.0
Libraries and Frameworks	fastText, Pandas, Numpy, Seaborn, Matplotlib, Nltk, Scikit-learn, Keras

C. Parameters Fine-tuning for Models

Table II lists the values of parameters set for these models.

TABLE II.   Parameters Settings

Model	Parameters	Value
BERT	Mode Transformer Batch Size Token Length Epoch Learning Rate Optimizer Loss	Bert-base-cased Hugging Face 16 160 10 0.0002 Adam Softmax
LSVM	Unigram, Bigram Trigram Kernel	 Linear
fastText	Unigram Bigram Trigram Epoch Learning Rate Loss	 10 0.01 Softmax
BLSTM	Epoch=10 Learning Rate Loss=softmax	10 0.01 Softmax
fastText-BLSTM	Epoch Learning Rate Loss	10 0.01 Softmax

D. Performance Measures

The models evaluated are based on accuracy, recall, precision and F1 score while the performance measures are calculated based on the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (TN) matrix. The following Table III lists the values of performance measures.

TABLE III. EVALUATION METRIC

Models	TP	FP	TN	FN
BERT	102913	11089	107371	12217
LSVM	95772	18687	105116	14015
fastText	107451	11680	98108	16351
BiLSTM	88764	25695	91100	28031
fastText-BiLSTM	104117	9885	106036	13552

The following are the equations to calculate the performance measures:

$$Precision = \frac{tp}{tp+fp} \qquad (8)$$

$$Recall = \frac{tp}{tp+fn} \qquad (9)$$

$$F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (10)$$

$$Accuracy\ \% = \frac{tp+tn}{tp+fp+tn+fn} * 100$$

The following Table IV and Table V show the performance measures and sentiment score of the models.

TABLE IV. MODELS PERFORMANCE MEASURES

Model	Accuracy %	Recall	Precision	F1
BERT	90%	0.91	0.90	0.90
LSVM	86 %	0.87	0.84	0.85
fastText	88%	0.87	0.89	0.88
BiLSTM	77%	0.75	0.77	0.76
fastText-BiLSTM	90%	0.87	0.91	0.89

TABLE V. SENTIMENT SCORE

Models	Positive	Negative
BERT	49.28%	50.71%
LSVM	46.99%	53.00%
fastText	52.99%	47.00%
BiLSTM	50.00%	50.00%
fastText-BiLSTM	50.37%	49.62%

The results of the accuracy are compared with other researchers' model results for the same datasets. The result showed BERT and fastText models' accuracy are 90% compared to other models. Table VI shows the comparison of model accuracy. Our experimental results performance measures show that fine-tuned Pre-trained BERT model scores highest F1 score of 0.90 and hybrid fastText-BiLSTM

model scores F1 score of 0.89 and other models scored less than 0.85. BERT shows the best result as a state-of-the-art model with respect to performance, accuracy, training and testing on customer reviews datasets.

TABLE VI. COMPARE MODELS ACCURACY SCORE

Et.al	Dataset	Model	Accuracy %
Ashok and Anandan	Twitter Customer Review	fastText Unigram Bigram Trigram	88.23 90.55 90.71
		LSVM Unigram Bigram Trigram	87.74 89.96 90.11
		SA-BLSTM	77.00
		fastText-BLSTM	90.00
		BERT	90.00
Geetika Gautam	Twitter Customer Review	SVM Max Entropy Naïve Bayes Semantic Analysis (WordNet)	85.50 83.80 88.20 89.90
Seyed-Ali Bahrainaian	Twitter Data on Smart phones	Hybrid Approach- Unigram, Naive Bayes, MaxEnt, SVM	89.78
Neethu M.S	Twitter data on electronic products	Naïve Bayes SVM Max Entropy Ensemble	89.50 90.00 90.00 90.00
Dhiraj Gurkhe	Twitter Data	Unigram Bigram Unigram-Bigram	81.20 15.00 67.50

V. CONCLUSION AND FUTURE WORK

The proposed BERT model outperforms in terms of accuracy and model performance compare to other models. The results of the fastText model showed low accuracy when unigram and bigram methods were used for training the model. The overall model training and data preparation tasks took less time for BERT model in comparison to others. This experiment reveals that the BERT model required more computational resources to train compared with other traditional models. The fastText model performed well with a standard server environment with minimal computational resources compare to other models. The fine-tuned BERT model simplifies the sentiment analysis tasks on large datasets.

A proposal for future research can be made to build transformer models for various domains, and framework for a continuous model trained to streamline data processing methods. In the future, more work will be done to conduct pre-training a BERT-base model on datasets of customer reviews for sentiment analysis and detecting emotions.

REFERENCES

[1] Connor Holmes, Daniel Mawhirter, Yuxiong He, Feng Yan, Bo Wu, "GRNN: Low-Latency and Scalable RNN Inference on GPUs", In Fourteenth Eurosys Conference (Eurosys '19) https://doi.org/10.1145/3302424.3303949, March, 2019.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakon Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, "Attention is all you need". 31st Conferencence on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, December, 2017.

[3] Jacob Devlin and Ming-Wei Chnag, Research Scientists Google AI Lanaguage, "https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html", November 2, 2018. Accesed on 10/30/2021.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, "BERT:Pre-training of Deep Birirectional Transformers for Lanaguage Understanding" Google AI Lanaguage, arXiv:1810.04805v2 [cs.CL] 24 May, 2019.

[5] Diego A. Andrade-Segarra1, Gabriel A. Leon-Paredes ´2, "Deep Learning-based Natural Language Processing Methods Comparison for Presumptive Detection of Cyberbullying in Social Networks", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 12, No. 5, May, 2021.

[6] Mohammed M.Abdelgwad, Taysir Hassan A Soliman, Ahmed I.Tabloba and Mohamed Fawzy Farghaly, "Arabic aspect based sentiment analysis using BERT", arXiv:2107.13290v2 [cs.CL] 28 September, 2021.

[7] Marco Pota, Mirko Ventura, Rosario Catelli, Massimo Esposito, "An Effective BERT-BasedPipleline for Twitter Sentiment Analysis: A Case Study in Italina", Sensors, MDPI Article, December 2020.

[8] M. A. H. Akhand, Arna Roy, Argha Chandra Dhar and Md Abdus Samad Kamal, " Recent Progress, Emerging Techniques, and Future Research Prospects of Bangla Machine Translation: A Systematic Review", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. September, 2021.

[9] Dat Quoc Nguyen, Thanh Vu and Anh Tuan Nguyen, "BERTweet: A pre-trained lanaguage model for English Tweets". Proceeding of the 2020 EMNLP, pages 9-14. November 16-20, 2020.

[10] Truong H. V Phan and Phuc Do,"BERT+vnKG: Using Deep Learning and Knowledge Graph to Improve Vietnamese Question Answering System", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 7, July, 2020.

[11] Akbar Karimi, Leonardo Rossi, Andrea Prati, "Improving BERT Performance for Aspect-Based Sentiment Analysis", Department of Engineering and Architecture, University of Parma. Italy. March, 2021.

[12] Luca Bacco, Andrea Cimino, Felice Dell'Orletta, and Mario Merone, "Explainable Sentiment Analysis: A Hierarchical Transformer-Based Extractive Summarization Approach", MDPI, Electronics 2021, 10, 2195. https://doi.org/10.3390/electronics10182195. September, 2021.

[13] Hu Xu, Lei Shu, Philip S. Yu, Bing Liu, "Understanding Pre-trained BERT for Aspect-based Sentiment Analysis", Department of Computer Science, University of Illinois at Chicago, IL, USA. Proceedings of the 28th International Conference on computational Linguistics, Barcelona, Spain, December, 2020.

[14] Guangyao Pang, Keda Lu, Xiaoying Zhu, Jie He, Zhiyi Mo, Zizhen Peng, and Baoxing Pu, "Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model", Hindawi, Wireless Communications and Mobile Computing, Volume 2021, Article ID 5534615, 13 pages. https://doi.org/10.1155/2021/5534615. 2021.

[15] Himanshu Batra, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal, "BERT-Based Sentiment Analysis: A Software Engineering Perspective", Indian Institute of Information Technology Allahabad, India, arXiv:2106.02581v3 [cs.CV], Jun, 2021.

[16] Prakhar Ganesh, Yao Chen, Xin Lou, Mohammad Ali Khan, Yin Yang, Hassan Sajjad, Preslav Nakov, Deming Chen, Marianne Winslett, "Compressing Large-Scale Transformer-Based Models: A Case Study on BERT", https://doi.org/10.1162/tacl a 00413. Transactions of the Association for Computational Linguistics, vol. 9, pp. 1061–1080, Jun, 2021.

[17] Mrityunjay Singh, Amit Kumar Jakhar, Shivam Pandey, "Sentiment analysis on the impact of coronavirus in social life using BERT model", Social Network Analysis and Mining, Licence to Springer-Verlag GmBH Austria, part os Springer Nature. February, 2021.

[18] M.P.Geetha and D.Karthika Renuka, "Improving the performance of aspect based sentiment analysis using fine-tuned Bert base Uncased model", International Journal of Intelligent Networks 2 (2021) 64–69. Jun, 2021.

[19] Fatima-ezzahra Lagrari and Youssfi Elkettani, "Customized BERT with Convolution Model: A New Heuristic Enabled Encoder for Twitter Sentiment Analysis", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 10, October, 2020.

[20] Majdi Baseiso and Saleh Alzahrani, "An Empirical Analysis of BERT Embedding for Autometed Eassay Scoring", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 10, October, 2020.

[21] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification", Facebook AI Research arXiv:1607.01759v3 [cs.CL], August, 2016.

[22] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov, "Enriching Word Vectorrs with Subword Information", Facebook AI Research, arXiv:1607.04606v2 [cs.CL] 19 Jun, 2017.

[23] Karthik Gopalakrishnan and Fathi M. Salem, "Sentiment Analaysis Using Simplified Long Short-term Memory Recurrent Neural Networks", Depatment of Electrical and Computer Engineering, Michigan State University USA. arXiv:2005.03993v1[cs.CL]. May, 2020.

[24] Surbhi Arora, "SVM: Difference between Linear and Non-Linear Models", 4 February, 2020.

[25] Alexandre Kowalczyk, "Support Vector Machine Succinctly", Syncfusion Inc, https://www.syncfusion.com/succinctly-free-ebooks/confirmation/support-vector-machines-succinctly October, 2017.