

Natural Language Processing of Movie Reviews to Detect the Sentiments using Novel Bidirectional Encoder Representation-BERT for Transformers over Support Vector Machine

Chikkili. Hema Kumar¹, R. Senthil Kumar²

¹Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode:602105.

²Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode:602105.

Abstract

Aim: The aim of the study is to detect sentiment analysis from the good ones and improve the false positivity rate by using the proposed Novel Bidirectional Encoder Representation for Transformers- (Novel BERT) over Support Vector Machine-SVM. **Materials and Methods:** Sample groups that are considered in the project can be classified into two, one for Novel BERT over SVM, which are tested using 0.80 for G-power to determine the sample size and for t-test analysis. 25000 IMDB movie reviews dataset that data collected from Twitter. **Results:** The automatic feature selection of the BERT algorithm splits the data with best fit, which has an average accuracy of 83.5%, which by far seems to be better than the SVM which gives average accuracy 75.3%. The significance is around 0.042 ($p < 0.05$) and therefore there is a statistically insignificant difference among the study group. **Conclusion:** BERT seems to be better in finding the Sentiment in IMDB movie dataset over the SVM algorithm.

Keywords: Novel BERT, Sentiment Analysis, Support Vector Machine, Transformer, Natural Language Processing, Tokenizer.

DOI: 10.47750/pnr.2022.13.S04.069

INTRODUCTION

In today's world, anyone can share their thoughts on social media. Aside from that, there's a movie review. Information regarding movie reviews is available on the internet thanks to technological advancements (C.-L. Liu et al. 2012). It is difficult to locate the correct information that the user needs. That can determine the value and quality of the information if it is processed. It will make it simple for people to determine the quality of a film by categorising movie reviews (Indrayuni 2018). The handling of negative words in reviews is a challenge in this movie review. The word negation refers to a statement that denies or reverses the value. 'Not' is the most common negation word. The choice of a negative term in a movie review may have an impact since it alters the sentimental value. Positive can turn by negative with negative words (B. Liu 2012). This sentiment analysis is one of the major applications that was devoted to movie reviews in English and was produced using a machine learning approach known as Support Vector Machine. Machine learning is popular and its applications are in a wide range of computer-related tasks, including text analysis, image processing, finance, search and recommendation engines, speech recognition, and so on (B. Liu 2012, 2020). In this scenario, machine learning is utilised to classify sentences based on good or negative user feedback (Yulietha et al. 2018).

There are around 150 google scholars and 190 IEEE research papers have been published over the past 6 years. The most cited article in IEEE is "Sentiment Analysis for Social Media using the SVM Classifier of Machine Learning". In the existing research they didn't identify the availability and efficient prediction of sentences. The main aim of the project is to identify the availability and efficiency of the proposed algorithm namely Novel BERT different methods and calculating the analysis of words function in reviews. Novel Bidirectional Encoder Representations-BERT from Transformers is a state of art machine learning approach for natural language processing. BERT was introduced by Jacob Devlin at Google in 2018. When it comes to data preprocessing, natural language processing (NLP) is one of the most difficult fields in artificial intelligence. It takes a long time

for the function to train successful NLP models, even after preprocessing, full processing and tokenizing text words (Panchal 2019; Rothman 2021a). The support vector machine-SVM algorithm approach was used to train the classifier model, coupled with feature selection utilising hybrid optimization. To extract previous polarity features and n-grams, The first do data pre-processing to reduce noisy data (Song, Hu, and Xie 2002). Then utilise Part of Speech (POS) tagging to extract prior polarity features and n-grams. After that, using Pointwise Mutual Information, the word polarity score is determined (PMI). Then, for feature optimization, a heuristic hybrid optimization strategy employing PSO and cuckoo search is applied, which aids in improving the model's performance ("Sentiment Analysis for Social Media Using SVM Classifier of Machine Learning" 2019).

Our team has extensive knowledge and research experience that has translate into high quality publications(Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021). In the existing research, they didn't identify the efficient algorithm, the system always uses a vectorization method, for deploying sentences to analyse the data without loss and better accuracy. It is a hybrid optimization technique, it presents the nearest value of expectation that leads to the data loss and accuracy of predicting the sentence of sentiment is slow. The main aim of our project is to identify the efficient algorithm for analysing the sentences by analysing the same set of data sets in different algorithms and calculating the accuracy.

Materials And Methods

The research work was performed in the Cloud Computing Laboratory Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS. The proposed work contains two groups. Group 1 is taken as the Novel BERT and group 2 as Support Vector Machine-SVM. The Bidirectional Encoder Representation for Transformation and Support Vector Machine were executed and evaluated a different number of times with a sample size of 20. A sample dataset is presented in Table 1. The experimental results are based on the analysis for pretest power which shows the high accuracy and high precision compared to the existing system. The minimum power analysis for G-Power calculation is fixed at 0.8 and the maximum accepted error is fixed at 0.5.

In Natural Language Processing, the Novel BERT algorithm is the most supervised learning algorithm which delivers high accuracy by using optimization procedures. There are a total of 25000 Comments that are considered as samples for a training data set which acts as the supervised learning method for the proposed Natural language processing with BERT for tokenization (Haughton et al. 2015).

Hugging Face Transformers Library Installation The location of the BERT model that I can use with TensorFlow was one of the problems that had the hard time resolving the problem. It is a high performance boosting framework based on the NLP algorithms used for deep learning tasks. In this algorithm it can easily create and reduce complexity. It handles large amounts of data, takes less training time and consumes low memory.

Testing setup for this proposed system used a Google Collab notebook with Jupyter enabled with GPU. Jupyter notebook is a software which is used for creating the Sentiment analysis with BERT model and SVM Hardware configuration for this proposed system is Intel core i5 8th gen processor and requires 4GB random access memory and 25 GB Solid state drive used. The configuration of the system is windows 11 operating system and jupyter notebook software and python programming language 3.9.8.

Testing procedure for Sentiment Analysis with Novel BERT

Novel Bidirectional Encoder Representation for Transformation, it is a bidirectional encoder to analyse the data from two sides to predict or analyse huge data within a short time period and Transformation is a machine learning package that includes models of NLP and algorithms to easily function. The both bidirectional and transformation to combine and represent the BERT model.

Step 1: Preprocessing / Text processing

In the preprocessing step it can remove the null/empty by data cleaning from the dataset which is not useful to analyse. It uses a normal NLP process to tokenize words to perform the operation.

Step 2: Feature Extraction

Feature extraction is used to extract the positive and negative opinion on the movie. This process of reduction by which an initial set of raw data is reduced to a manageable group for processing.

Step 3: Training

After the feature extraction step, the training process is done. Training step involves implementation of BERT to predict the sentiment analysis. This analyses the dataset into two parts in that train contains 28000 to train with the algorithm.

Step 4: Test Classifier

In the test classifier can import different and more datasets to test the accuracy of the classifier and recall . Test contains 13000 comments for testing the data set into to get recall and f1-score.

Step 5: Testing

Testing process is used to detect if the comments in the dataset are positive or negative. In this step the classifier shows the result of spam if it shows positive and negative .It analyses the words with negative or positive then it predicts the sentence.

Step 6: Finding the predictions

Using the natural language processing with BERT algorithms to predict the text contains opinions or emotions in positive or negative . It predicts the sentence to perform the operation and deceit the analysis of the given sentence.

Sentiment Analysis using BERT and Tensorflow

BERT: Bidirectional Encoder Representation for Transformation Algorithm it was developed in 2018 by google.

Input: IMDB dataset, prediction statement.

```
Model← TFBERT_ForSequenceClassification.from_pretrained()
```

```
Tokenizer← BERT_Tokenizer.from_pretrained() \\ Tokenize the sentence to predict with BERT
```

```
For i in dataset :
```

```
    train , test← tf.keras.preprocessing\\ it processing the data
```

```
End for
```

```
For i to train,test:
```

```
    train_feat , train_lab \\ code converts our train Dataset object to train pandas data frame
```

```
    test_feat , test_lab \\ code converts our test Dataset object to test pandas data frame
```

```
End for
```

```
For e in tokenizer:
```

```
    input_dirTokenizer.encode()
```

```
    max_length=length           \\if len(s)>max_length
```

```
    features.append()           \\ appending the input sequence
```

```
End for
```

```
\\ Configuring the Novel BERT model and Fine- tuning
```

```
Model.compile()                \\ it measures the accuracy and loss of algorithm
```

```
Model.fit(train_data,epochs)    \\ epochs indicates no.of passes in training dataset
```

```
Prediction_sentence()          \\ input of the sentences to predict
```

```
For predict ← range(predict_sentence):
```

```
    prediction_sentence
```

```
End for
```

Output: Accuracy , predicted sentence is positive or negative

End

SVM- Support Vector Machine

SVM-Support Vector Machine is a prominent Supervised Learning technique that may be used to solve both classification and regression problems. Supervised learning models with accompanying learning algorithms that evaluate data for classification and regression analysis are known as support-vector machines. The SVM is to find the optimum line or decision boundary for categorising n-dimensional space into classes so that additional data points can be readily placed in the correct category in the future. A hyperplane is the name for the optimal choice boundary. In statistical learning theory, one property of SVM is the intrinsic risk minimization. SVM are a class of supervised learning methods for classification, regression, and outlier detection. They are effective in high-dimensional spaces. When the number of dimensions exceeds the number of samples, the method is still effective.

SVM Algorithm

Input: IMDB dataset, prediction statement.

```
Model← SVM_ForSequenceClassification.from_pretrained()
Tokenizer←SVM_Tokenizer.from_pretrained() \\ Tokenize the sentence to predict with SVM
For i in dataset:
train ,test← tf.keras.preprocessing          \\it processing the data
End for
For i to train,test:
    train_feat, train_lab \\ code converts our train Dataset object to train pandas data frame
    test_feat, test_lab \\ code converts our test Dataset object to test pandas data frame
End for
Vectorization ← Tfidfvectorizer()          \\ train ,test_vectorizer for fit transform
                                          \\ Creating a Linear SVM model
Classifier_linear=SVM.svc()
Time_linear_train()                        \\ time prediction
Classifier_linear.fit()                     \\ prediction_linear with predict of time
Model.compile()                             \\ it measures the accuracy and loss of algorithm
Model.fit(train_data,epochs)
                                          \\ epochs indicates no.of passes in training dataset
Prediction_sentence()                       \\input of the sentences to predict
For predict ← range(predict_sentence):
    prediction_sentence
End for
Output: Accuracy , predicted sentence is positive or negative
End
```

Dataset is collected from IMDB Dataset. In the dataset 70% is used for training and 30% used for testing. The BERT and SVM algorithms were evaluated with respect to training, and tests were conducted with the required parameters to improve the accuracy percentage.

Statistical Analysis

Statistical software used in the study is IBM SPSS version 26 . The independent sample T-test calculation for analysing equal variance, standard error, and levene's test are evaluated. Attributes like comment ID, Author, Data and type of comments form the independent variables and Comments, Detection and class are dependent variables. Independent sample T-test has been carried out for evaluating the accuracy (Singh, Jakhar, and Pandey 2021).

Results

In this proposed system it was observed that BERT uses optimization procedures to deliver high accuracy and no data loss with bidirectional. Where Table 2 shows the data present in IMDB dataset that contain a positive and negative opinion information count, Fig. 2 represents a count of positive and negative opinions. Table 3 shows the statistical calculation such as mean, standard deviation and standard error mean for NBERT and SVM algorithms. The mean, standard deviation and standard error mean for BERT are 83.5, 8.3021, 3.029 respectively. The mean, standard deviation and standard error mean for SVM are 75.3, 9.3446,4.496 respectively. It is inferred that the mean accuracy for NBERT is 83.5 which is greater than the mean accuracy of comparison architecture which is 75.3. Moreover, the mean accuracy value of Novel Bidirectional Encoder Representation for Transformation is 83.5 which seems to be superior to the Support Vector Machine. Fig. 1 explains the architecture of NBERT with bidirectional analysis of the dataset. In Table 4, it was observed that the Levens test for equality of variance and its significance for BERT is 0.747 and 0.302, respectively and standard error difference and confidence interval are lower than SVM classifiers. The mean accuracy is depicted in Fig. 3. For the supplied dataset of Sentiment Analysis on IMDB Dataset, BERT appears to be superior to movie review comments.

Discussion

The proposed system provides better sentiment analysis using a BERT with a count vectorizer with over 95.1% accuracy compared to SVM for improved Classification. In industry sentiment analysis with BERT and SVM performed the best among the model combinations created on all datasets, according to the results of the testing. The mean, standard deviation and standard error mean for Bidirectional Encoder Representation for Transformation deployment are 83.5, 8.3021, 3.029 respectively. The mean, standard deviation and standard error mean for SVM are 75.3, 9.3446, 4.496 respectively. It is inferred that the mean accuracy for NBERT is 83.5 which is greater than the mean accuracy of comparison architecture which is 75.3. A systematic review on Sentiment Analysis on movie reviews presents around 45 studies, the analysis of various papers shows that BERT is the most used algorithm for text classification (Rothman 2021b). Since most of the movie reviews sentiment are related to text links, the algorithm provides the best accuracy for finding positive and negative comments (Singh, Jakhar, and Pandey 2021). Sentiment analysis with the BERT algorithm has more accuracy compared to the SVM. The Support Vector Machine was used for classification. A learning machine for two-group classification issues is the support-vector machines (Moore and Wang 2018). It's used to categorise writings as either positive or negative. Because of its advantages, such as its ability to handle huge features, SVM performs well for text classification (Joachims 2012). Another benefit is that SVM is robust in the presence of a sparse set of samples, as well as the fact that most problems are linearly separable. In past sentiment analysis studies, Support Vector Machines have demonstrated promising results (Onal, Ertugrul, and Cakici 2014).

In industry sentiment analysis data, according to the results of the testing. BiLL is a bidirectional LSTM model with knowledge distilled from bidirectional encoder representation for transformation (Chebotar and Waters 2016). They combine multiple heuristics with rule-based data augmentation. The contribution of data augmentation is not addressed in this situation because it is employed as a complimentary tool for the task. AL-BERT is a simplified variant of BERT with fewer parameters than the regular BERT design. Factorised embedding parameterization and cross-layer parameter sharing are two strategies for parameter reduction (Chi et al. 2021). These methods lower BERT's memory use and reference time. BERT is reduced to a tiny convolutional neural network model for a specific job (S. Liu, Tao, and Feng 2019). They produce sentences in the style of training data using GPT-2, a big language model. The training set is supplemented by the synthetic sentences (Azunre 2021).

The limitation of the study is scaled if the analysis is of various words, and the classifier runs on each sentence. In such scenarios, the accuracy is not particularly high, and it can potentially fail in unexpected ways. Further, this research work can be improved by deploying a model that predicts more appropriate words in less time so that wait will be less and it can be easily manageable and scalable as in this research.

Conclusion

The Novel Bidirectional Encoder Representation for Transformers-BERT algorithm identifies the sentiment analysis of movies with better accuracy of 83.5% compared to Support Vector Machine-SVM algorithm with 75.3%. BERT algorithm that helps to function in both directions to predict the analysis on dataset with bidirectional encoder. Support Vector machine that functions on vectorization process to predict the sentiment analysis, so it takes a longer time and loss of data than BERT.

DECLARATIONS

Conflict of Interests

No conflict of interest in this manuscript.

Author Contribution

Author CHK was involved in data collection, data analysis, and manuscript writing. Author UKS was involved in conceptualization, guidance and critical review of manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organisations for providing financial support that enabled us to complete the study.

1. STEP UP solutions, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

References

1. Adhinarayanan, Rajesh, Aravindh Ramakrishnan, Gopal Kaliyaperumal, Melvin Victor De Poures, Rajesh Kumar Babu, and Damodharan Dillikannan. 2020. "Comparative Analysis on the Effect of 1-Decanol and Di-N-Butyl Ether as Additive with diesel/LDPE Blends in Compression Ignition Engine." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
2. Arun Prakash, V. R., J. Francis Xavier, G. Ramesh, T. Maridurai, K. Siva Kumar, and R. Blessing Sam Raj. 2020. "Mechanical, Thermal and Fatigue Behaviour of Surface-Treated Novel Caryota Urens Fibre-reinforced Epoxy Composite." *Biomass Conversion and Biorefinery*, August. <https://doi.org/10.1007/s13399-020-00938-0>.
3. Aurtherson, P. Babu, Bhanu Teja Nalla, Karthikeyan Srinivasan, Kulmani Mehar, and Yuvarajan Devarajan. 2021. "Biofuel Production from Novel Prunus Domestica Kernel Oil: Process Optimization Technique." *Biomass Conversion and Biorefinery*, May. <https://doi.org/10.1007/s13399-021-01551-5>.
4. Azunre, Paul. 2021. *Transfer Learning for Natural Language Processing*. Simon and Schuster.
5. Bhansali, Karan J., Kamlesh R. Balinge, Subodh U. Raut, Shubham A. Deshmukh, M. Senthil Kumar, C. Ramesh Kumar, and Pundlik R. Bhagat. 2021. "Visible Light Assisted Sulfonic Acid-Functionalized Porphyrin Comprising Benzimidazolium Moiety for Photocatalytic Transesterification of Castor Oil." *Fuel* 304 (November): 121490.
6. Chebotar, Yevgen, and Austin Waters. 2016. "Distilling Knowledge from Ensembles of Neural Networks for Speech Recognition." *Interspeech 2016*. <https://doi.org/10.21437/interspeech.2016-1190>.
7. Chi, Po-Han, Pei-Hung Chung, Tsung-Han Wu, Chun-Cheng Hsieh, Yen-Hao Chen, Shang-Wen Li, and Hung-Yi Lee. 2021. "Audio Albert: A Lite Bert for Self-Supervised Learning of Audio Representation." *2021 IEEE Spoken Language Technology Workshop (SLT)*. <https://doi.org/10.1109/slt48900.2021.9383575>.
8. Deepanraj, B., N. Senthilkumar, D. Mala, and A. Sathiamourthy. 2021. "Cashew Nut Shell Liquid as Alternate Fuel for CI Engine—optimization Approach for Performance Improvement." *Biomass Conversion and Biorefinery*, February. <https://doi.org/10.1007/s13399-021-01312-4>.
9. Haughton, Dominique, Mark-David McLaughlin, Kevin Mentzer, and Changan Zhang. 2015. *Movie Analytics: A Hollywood Introduction to Big Data*. Springer.
10. Indrayuni, Elly. 2018. "KOMPARASI ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISA SENTIMEN REVIEW FILM." *Jurnal Pilar Nusa Mandiri*. <https://doi.org/10.33480/pilar.v14i2.918>.
11. Jayanth, Bellappu Venkat, Melvin Victor Depoures, Gopal Kaliyaperumal, Damodharan Dillikannan, Dilipsingh Jawahar, Kumaran Palani, and Ganesha Prasad Meravanigee Shivappa. 2021. "A Comprehensive Study on the Effects of Multiple Injection Strategies and Exhaust Gas Recirculation on Diesel Engine Characteristics That Utilize Waste High Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
12. Joachims, Thorsten. 2012. *Learning to Classify Text Using Support Vector Machines*. Springer Science & Business Media.
13. Kamath, Manjunath, Subha Krishna Rao, Jaison, Sridhar, Kasthuri, Gopinath, Sivaperumal, and Shantanu Patil. 2020. "Melatonin Delivery from PCL Scaffold Enhances Glycosaminoglycans Deposition in Human Chondrocytes – Bioactive Scaffold Model for Cartilage Regeneration." *Process Biochemistry* 99 (December): 36–47.
14. Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
15. ———. 2020. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
16. Liu, Chien-Liang, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou. 2012. "Movie Rating and Review Summarization in Mobile Environment." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. <https://doi.org/10.1109/tsmcc.2011.2136334>.
17. Liu, Songsong, Haijun Tao, and Shiling Feng. 2019. "Text Classification Research Based on Bert Model and Bayesian Network." *2019 Chinese Automation Congress (CAC)*. <https://doi.org/10.1109/cac48633.2019.8996183>.
18. Moore, Candace, and David Wang. 2018. "Support Vector Machine (machine Learning)." *Radiopaedia.org*. <https://doi.org/10.53347/rid-61710>.
19. Onal, Itir, Ali Mert Ertugrul, and Rukun Cakici. 2014. "Effect of Using Regression on Class Confidence Scores in Sentiment Analysis of Twitter Data." *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. <https://doi.org/10.3115/v1/w14-2622>.
20. Panchal, Akash. 2019. "NLP — Text Summarization Using NLTK: TF-IDF Algorithm." *Towards Data Science*. June 10, 2019. <https://towardsdatascience.com/text-summarization-using-tf-idf-e64a0644ace3>.
21. Rajasekaran, S., D. Damodharan, K. Gopal, B. Rajesh Kumar, and Melvin Victor De Poures. 2020. "Collective Influence of 1-Decanol Addition, Injection Pressure and EGR on Diesel Engine Characteristics Fueled with diesel/LDPE Oil Blends." *Fuel* 277 (October): 118166.
22. Rajesh, A., K. Gopal, De Poures Melvin Victor, B. Rajesh Kumar, A. P. Sathiyagnanam, and D. Damodharan. 2020. "Effect of Anisole Addition to Waste Cooking Oil Methyl Ester on Combustion, Emission and Performance Characteristics of a DI Diesel Engine without Any Modifications." *Fuel* 278 (October): 118315.
23. Raju, P., K. Raja, K. Lingadurai, T. Maridurai, and S. C. Prasanna. 2021. "Glass/Caryota Urens Hybridized Fibre-Reinforced nanoclay/SiC Toughened Epoxy Hybrid Composite: Mechanical, Drop Load Impact, Hydrophobicity and Fatigue Behaviour." *Biomass Conversion and Biorefinery*, March. <https://doi.org/10.1007/s13399-021-01427-8>.
24. Rothman, Denis. 2021a. *Transformers for Natural Language Processing: Build Innovative Deep Neural Network Architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and More*. Packt Publishing Ltd.
25. ———. 2021b. *Transformers for Natural Language Processing: Build Innovative Deep Neural Network Architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and More*. Packt Publishing Ltd.
26. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). <https://doi.org/10.1002/ep.13696>.
27. "Sentiment Analysis for Social Media Using SVM Classifier of Machine Learning." 2019. *Special Issue Report. International Whaling Commission* 8 (9S4): 39–47.
28. Shanmugam, Rajasekaran, Damodharan Dillikannan, Gopal Kaliyaperumal, Melvin Victor De Poures, and Rajesh Kumar Babu. 2021. "A Comprehensive Study on the Effects of 1-Decanol, Compression Ratio and Exhaust Gas Recirculation on Diesel Engine Characteristics Powered with Low Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 43 (23): 3064–81.
29. Singh, Mrityunjay, Amit Kumar Jakhar, and Shivam Pandey. 2021. "Sentiment Analysis on the Impact of Coronavirus in Social Life Using the BERT Model." *Social Network Analysis and Mining* 11 (1): 33.
30. Song, Qing, Wenjie Hu, and Wenfang Xie. 2002. "Robust Support Vector Machine with Bullet Hole Image Classification." *IEEE*

- Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews: A Publication of the IEEE Systems, Man, and Cybernetics Society* 32 (4): 440–48.
31. Sudhakar, M. P., Merlyn Ravel, and K. Perumal. 2021. "Pretreatment and Process Optimization of Bioethanol Production from Spent Biomass of *Ganoderma Lucidum* Using *Saccharomyces Cerevisiae*." *Fuel* 306 (December): 121680.
 32. Yulietha, I. M., S. A. Faraby, Adiwijaya, and W. C. Widyaningtyas. 2018. "An Implementation of Support Vector Machine on Sentiment Classification of Movie Reviews." *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/971/1/012056>.

TABLES AND FIGURES

Table 1. The data that was collected from twitter about movie reviews in text normalisation with the public opinion about it in positive or negative way of comments. This dataset name is IMDB_dataset.

S.No	Data_column	Label_coloum
0	One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked.	positive
1	The filming technique is very unassuming- very old-time-BBC fashion and gives a comforting, and sometimes discomforting, sense of realism to the entire piece.	positive
2	Basically there's a family where a little boy thinks there's a zombie in his closet & his parents are fighting all the time.	Negative
3	For a long time, 'The Menagerie' was my favori..	positive
4	A truly frightening film. Feels as if it were	Negative

Table 2. The Count of positive sentences on movie opinion with text representation or negative sentence on movie opinion with text representation reviews in the dataset. The IMDB dataset collects data from Twitter analysis.

In IMDB- Dataset comments	Positive	Negative
Count	27000	23000

Table 3. Group statistical analysis of Bidirectional Encoder Representation with mean value of 83.5 and Support vector Machine 75.3 and similarly the results of Standard Deviation and Standard Error Mean are given for both BERT and SVM.

	Algorithm	N	Mean	Std. Deviation	Std.Error Mean
Accuracy	BERT	20	83.5	8.3021	3.029
	SVM	20	75.3	9.3446	4.496

Table 4. Independent Sample T-test Results with confidence interval of 95% and level of significance greater than 0.05 (BERT appears to perform significantly better than SVM with the value of $p > 0.001$)

	Levene's Test for Equality of Variance	Levene's Test for Equality of Variance									
		Equal Variances	F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
										Lower	Upper
Accuracy	Assumed	.004	.042	-.397	3	.001	5.500	5.614	6.976	9.176	
	Not Assumed			-.397	7.860	.001	5.5040	5.6056	6.976	9.187	

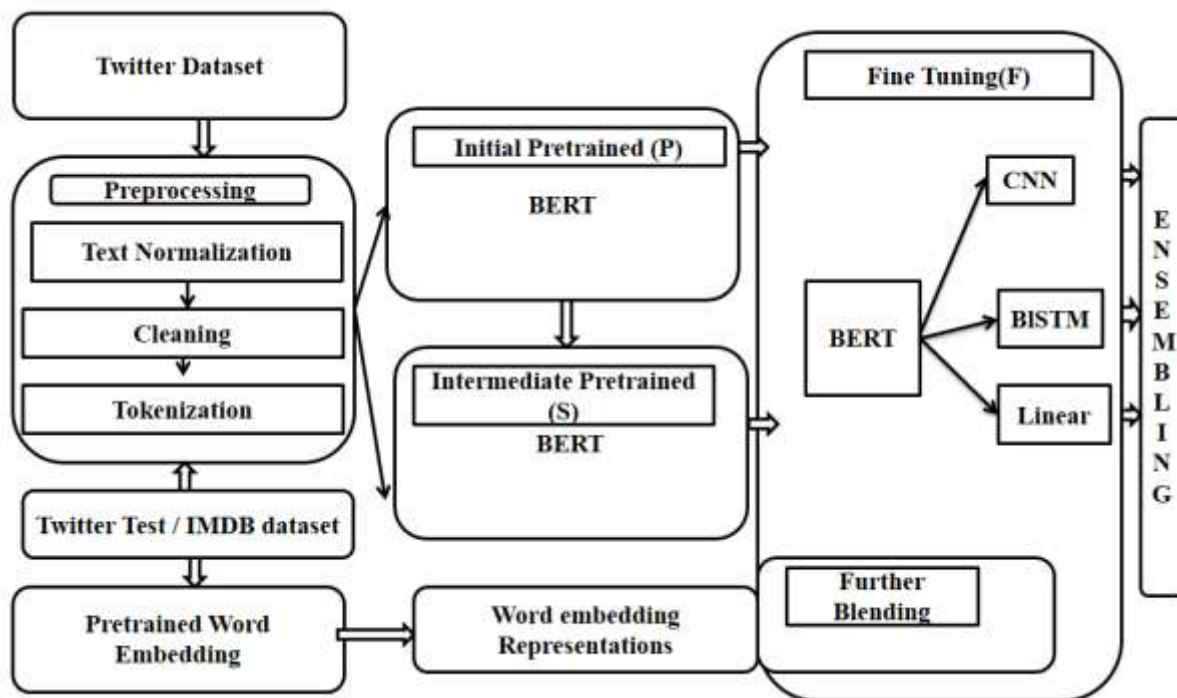


Fig. 1. Architecture for Sentiment analysis with Bidirectional Encoder Representation for Transformation on IMDB movie reviews. This imports the text normalisation process to analyse the data and convert into word embedding to predict the analysis to produce the sentiment on given sentences in the architecture.

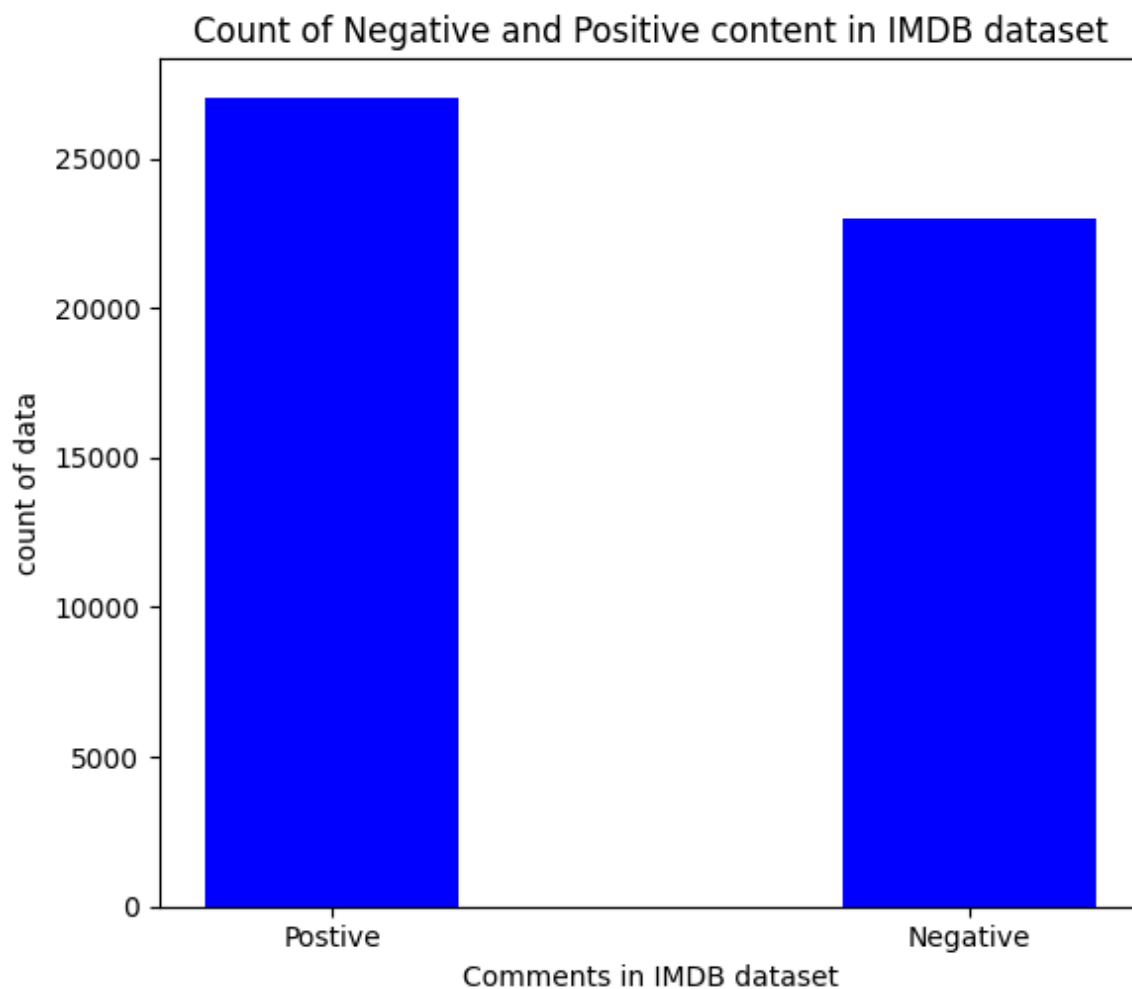


Fig. 2. Dataset that contains text forms with positive sentences on movie opinion with text representation and negative sentences on movie opinion with text representation. That contains 27000 positive reviews and 23000 negative reviews on IMDB.

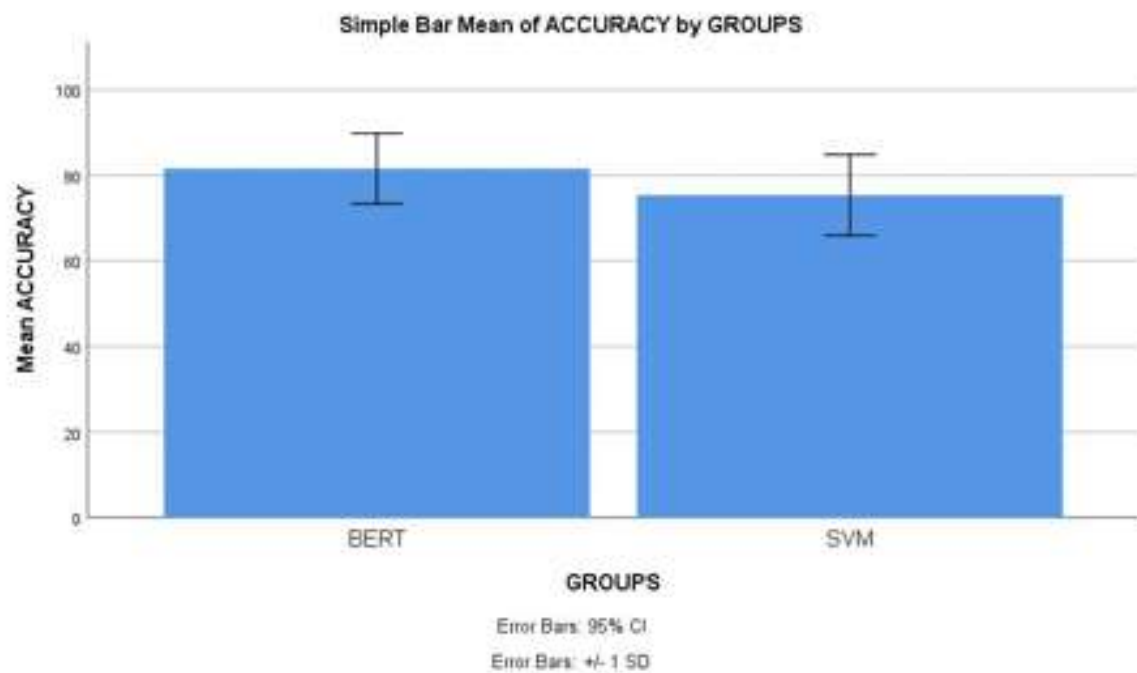


Fig. 3. Bar graph analysis of Novel Bidirectional Encoder Representation for Transformation architecture and Support vector Machine architecture. Graphical representation shows the mean Accuracy of 83.5% and 75% for the proposed NBERT and SVM respectively. X-axis : BERT vs SVM, Y-axis : Mean Accuracy \pm 1 SD.