Origin Article

Analysing the Sentiments in e-newspaper Contents using Novel Bidirectional Encoder Representation for Transformation - BERT over Linear Regression Algorithm

Chikkili. Hema Kumar[1], Malathi.K[2*]

[1]Research 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.

[2*]Project 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-BERT over Linear Regression Algorithm.

Materials and Methods: Sample groups that are considered in the project can be classified into two, one for Novel BERT over Linear Regression, which are tested using 0.80 for G-power to determine the sample size 20 and for t-test analysis. 23000 BBC e-news dataset that collected data from NewsBrief and MediSys.

Results: The automatic feature selection of the BERT algorithm splits the data with best fit, which has an average accuracy of 83.50%, which by far seems to be better than the Linear regression which gives around 77.80%. The significance is around 0.039 ($p<0.05$) and therefore there is a statistical insignificant difference among the study group.

Conclusion: Novel BERT seems to be better in finding the Sentiment in e-newspaper content of BBC e-news dataset over the Linear Regression Algorithm.

Keywords: Sentiment Analysis, Novel BERT, Predictive, Linear Regression, Transformers, Newspapers.

DOI: 10.47750/pnr.2022.13.S04.004

INTRODUCTION

The computational study of people's opinions, attitudes, and feelings toward an entity is known as sentiment analysis (SA). Individuals, events, or concepts can all be represented by the entity. Reviews are more likely to discuss these subjects. The terms SA can be used interchangeably. They are expressing a shared meaning. The applications of opinion mining extracts and analyses people's views on a subject, whereas Predictive Analysis recognises and analyses the sentiment expressed in a document (Singh, Jakhar, and Pandey 2021). NewsBrief and MedISys are BBC systems that categorise news into one or more of several hundred subject domain groups, such as natural catastrophes, security, finance, nuclear issues, various diseases, organisations, countries, regions, specific conflicts (Balahur, Steinberger, et al. 2009). This negative spike occurs regardless of an entity's role in the events; for example, even if a person is seeking to act favourably in the event, the sentiment value toward that person may be negative Newspapers (Balahur, Steinberger, et al. 2009). As a result, in our recent opinion mining efforts, concentrated on examining smaller and bigger word windows around entities, and attempted to extract positive and negative sentiment from good and bad news ("IAPWNC 2009: Intelligent Analysis and Processing of Web News Content" 2009; Gupta, Abd El-Latif, and Perakovic 2021).

In order to shorten the time period to predictive analysis on sentences, lower the cost of developing and deploying containerized applications and speed up procedures, paradigms are being investigated. There are around 150 IEEE papers and 110 google scholar papers have been published over the past 5 years. The most cited article is "Prediction using Linear Regression Algorithm based on Sentiment Analysis". In the existing research they didn't identify the availability and efficient predictive analysis of sentences. The main aim of our project is to identify the availability and efficient algorithm for deploying applications by deploying the same set of algorithms in different methods and calculating the accuracy. NBERT (Bidirectional Encoder Representations

from Transformers) is a state of art machine learning approach for natural language processing. BERT was developed by Jacob Devlin and google colleagues 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 to train successful NLP models, even after preprocessing and tokenizing text sentences (Panchal 2019; Rothman 2021). Word embedding for newspapers is a prominent NLP-Natural Language Processing method for learning low-dimensional vector representations of words from predictive documents. Word embedding algorithms like BERT are useful because they can capture syntactic and semantic word associations. The profusion of technology and research drives information needs(Thomas 2020). The handling of negative words in reviews is a challenge in this e-news review. The word negation refers to a statement that denies or reverses the value. 'Not' is the most common negation word(Ghasiya and Okamura 2021). The choice of a negative term in a movie review may have an impact since it alters the sentimental value. Positives can turn negative with negative words.Linear Regression, a linear model was developed to determine the sentiment analysis on an e-newspaper. The linear model, despite its simplicity, performs admirably. However, a careful selection of the model's many "ingredients", such as feature selection and regularisation term, could improve its accuracy even more. In this paper, look at Boolean Information Gain can be used to choose Unigram features and compare the performance of the resulting Boolean and Multinomial models of the newspaper (Gelman and Hill, n.d.). The value of having numerous characteristics is more substantial than the difference between the Boolean and Multinomial models, according to our findings.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 that the system always uses a vectorization method for newspaper comments. It is a linear technique, the efficient algorithm for deploying sentences to analyse the data without loss. It presents the accurate value of expectation that leads to the loss of data and accuracy of predicting the sentence of sentiment. Also only checks the data in a linear way. 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 Bidirectional Encoder Representation for Transformation and group 2 as Linear Regression (Deepak et al. 2021). The Bidirectional Encoder Representation for Transformation and Linear Regression 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. Same set of 20 containerized applications are used to calculate the deployment time, management and scalability of the applications for each architecture to get the accuracy of each architecture.

In Natural Language Processing, the 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 .

Hugging Face Transformers Library of the BERT model 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 Linear Regression 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 BERT-Bidirectional Encoder Representation for Transformation, it is a bidirectional encoder to analyse the data from two sides to analyse huge data within a short period and Transformation package that includes models of NLP and algorithms to easily function.

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 14000 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 9000 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 sentance.

Step 6: Finding the predictions

Using the natural language processing with Novel 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 Novel BERT and Tensorflow

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

Input: BBC e-news 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 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_sentance):
        prediction_sentence
End for
```

Output: Accuracy , predicted sentence is positive or negative

End

LRA-Linear Regression Algorithm

Linear regression is a statistical model for predictive analysis. Sales, salary, age, product price, and other continuous/real or numeric variables are predicted using linear regression.The term "linear regression" refers to a relationship between a dependent (y) and one or more independent (y) variables. Because linear regression reveals a linear relationship, it determines how the value of the dependent variable changes as the value of the independent variable changes.They are mainly two types : one is Simple Linear Regression, it is a Linear Regression approach that uses a single independent variable to predict the value of a numerical dependent variable.and other one is Multiple Linear Regression, it is a Linear Regression approach that uses more than one independent variable to predict the value of a numerical dependent variable.

Algorithm

LRA- Linear Regression Algorithm

Input : BBC e-news dataset, prediction statement.

```
def process_news():
        tokenizer <-- newsTokenizer()          \\  instantiate tpkenizer
        for words in news:
                stem_word=stemmer.stem()
                news_stem.append()
return news
def gradientDescent():
        x,y                                    \\ x of matrix and Y dimensions
        for i<- 0 to n do:
                h=sigmoid()                    \\ sigmoid function
                z=np.dot()                     \\ dot product
return
def extract_features():
        For i <-train,test do:
                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
Time_linear_train()                        \\ time prediction
Classifier_linear.fit()  \\ Predictive_linear with predict of time
Model.compile() \\ it measures the accuracy and loss of algorithm
Model.fit(train_data,epochs)
return
def predict_news():
        test=np.array()
        accuracy=np.sum()
Print ← sentence
return accuracy
```

Output: Accuracy , Predictive sentence is positive or negative

END

Dataset is collected from BBC e-news dataset from newspaper. In the dataset 60% is used for training and 40% used for testing. The NBERT and Linear Regression Algorithm 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(Rothman 2021) .

RESULTS

In this proposed system it was observed that BERT uses optimization procedures to deliver high accuracy and no data loss with bidirection. BERT uses optimization procedures to deliver high accuracy. Where Table 1 shows the data present in BBC e-news that contain a positive and negative opinon information count, Fig. 1 represents a count of positive and negative opinion. Table 2 shows the statistical calculation such as mean, standard deviation and standard error mean for BERT andLinear Regression Algorithm algorithms. The mean, standard deviation and standard error mean for BERT are 83.50, 8.258, 3.693 respectively. The mean, standard deviation and standard error mean for Linear Regression Algorithm are 77.80, 8.167, 3.653 respectively. It is inferred that the deviation for T-test is far lesser than the comparison algorithm. Moreover, the accuracy value of BERT is 83.5 while the loss is around 13.5, which seems to be superior to the linear regression algorithm classifier. Figure 1 explains the architecture of NBERT with bidirectional analysis of the dataset, and the dataset is presented in Fig. 2. In Table 3, it was observed that the Levens test for equality of variance and its significance for BERT is 1.778 and 0.329, respectively and standard error difference and confidence interval are lower than Linear Regression Algorithm on predictive analysis. The mean accuracy and mean loss graph is depicted in Fig. 3. BERT seems to appear better for the given dataset of Sentiment Analysis on BBC e-newspaper dataset over comments of movie reviews.

DISCUSSION

The proposed system provides better sentiment analysis using a BERT with a count vectorizer with over 83.50% accuracy better than compared to Linear Regression Algorithm for improved Classification. In industry sentiment analysis with BERT and Linear Regression Algorithm 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 BERT are 83.50, 8.258, 3.693 respectively. The mean, standard deviation and standard error mean for Linear Regression Algorithm are 77.80, 8.167, 3.653 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 77.80.

A systematic review on Sentiment Analysis on movie reviews presents around 35 studies, the analysis of various papers shows that BERT is the most used algorithm for text classification ("IAPWNC 2009: Intelligent Analysis and Processing of Web News Content" 2009). 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 Linear Regression (Gelman and Hill, n.d.). That wanted to make sure that: a) estimate the opinion on the quote's target-by computing the opinion in windows of words between the entity's mentions, and b) eliminate the bad versus good news content-by removing those words that are both sentiment-bearing words and words that are part of BBC category definitions, now referred to as category words (Balahur, Boldrini, et al. 2009). News article with a sentiment score of 0 were deemed neutral, while those with a score of +1 were considered positive, and those with a sentiment value of -1 were considered negative. Table 1 summarises the experiment's findings.

Preprocessing is a crucial step for cleaning text and removing irregularities so that the data may be used more efficiently in text mining or sentiment analysis (Nhlabano and Lutu 2018). The entire preprocessing task was completed with the Rapid miner tool, which includes a large number of preprocessing operators. Using the "Tokenize" operator, the initial preprocessing task was to tokenize the text in news articles into a set of tokens (Udhayakumar et al. 2021). Tokenizing is the process of breaking down a series of sentences into individual components, such as words, phrases, or symbols, known as tokens (Pahwa, Taruna, and Kasliwal 2018). Linear regression is a statistical procedure that uses X features to predict a Y result. The data sets are analysed using machine learning to see if there is a correlation. The relationships are then plotted on an X/Y axis with a straight line connecting them to anticipate future relationships (Gelman and Hill, n.d.). BERT-AI is a BERT version with fewer parameters than the standard BERT design. Parameter reduction strategies include factored embedding parameterization and cross-layer parameter sharing (Chi et al. 2021). BERT is reduced to a compact convolutional neural network model for a specific job using these methods, which reduces memory usage and reference time (Liu, Tao, and Feng 2019). Using GPT-2, a large language model, they generate phrases in the

style of training data. The synthetic sentences are used to enrich the training set (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 BERT-Bidirectional Encoder Repsornation for Transformers algorithm identifies the sentiment analysis of movies with better accuracy of 83.5% compared to Linear Regression algorithm with 77.8%. BERT algorithm that helps to function in both directions to predict the analysis on dataset with bidirectional encoder.Linear Regression that functions in a linear way 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, manuscript writing. Author KM was involved in conceptualization, guidance and critical review of manuscript.

Acknowledgements
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, Melvinvíctor 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. Balahur, Alexandra, Ester Boldrini, Andrés Montoyo, and Patricio Martínez-Barco. 2009. "Opinion and Generic Question Answering Systems." *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers on - ACL-IJCNLP '09*. https://doi.org/10.3115/1667583.1667632.
6. Balahur, Alexandra, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen, and Mijail Kabadjov. 2009. "Opinion Mining on Newspaper Quotations." *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. https://doi.org/10.1109/wi-iat.2009.340.
7. 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.
8. 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.
9. Deepak, P., Tanmoy Chakraborty, Cheng Long, and Kumar G. Santhosh. 2021. *Data Science for Fake News: Surveys and Perspectives*. Springer Nature.

10. 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.
11. Gelman, Andrew, and Jennifer Hill. n.d. "Linear Regression: The Basics." *Data Analysis Using Regression and Multilevel/Hierarchical Models*. https://doi.org/10.1017/cbo9780511790942.005.
12. Ghasiya, Piyush, and Koji Okamura. 2021. "Investigating COVID-19 News Across Four Nations: A Topic Modeling and Sentiment Analysis Approach." *IEEE Access : Practical Innovations, Open Solutions* 9 (March): 36645–56.
13. Gupta, Brij, Ahmed A. Abd El-Latif, and Dragan Perakovic. 2021. *Data Mining Approaches for Big Data and Sentiment Analysis in Social Media*.
14. "IAPWNC 2009: Intelligent Analysis and Processing of Web News Content." 2009. *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. https://doi.org/10.1109/wi-iat.2009.413.
15. 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.
16. 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.
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. Nhlabano, V. V., and P. E. N. Lutu. 2018. "Impact of Text Pre-Processing on the Performance of Sentiment Analysis Models for Social Media Data." *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*. https://doi.org/10.1109/icabcd.2018.8465135.
19. Pahwa, Bhumika, S. Taruna, and Neeti Kasliwal. 2018. "Sentiment Analysis- Strategy for Text Pre-Processing." *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2018916865.
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. 2021. *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. 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.
26. 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.
27. 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.
28. 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.
29. Thomas, Alex. 2020. *Natural Language Processing with Spark NLP: Learning to Understand Text at Scale*. O'Reilly Media.
30. Udhayakumar, S., J. Silviya Nancy, D. UmaNandhini, P. Ashwin, and R. Ganesh. 2021. "Context Aware Text Classification and Recommendation Model for Toxic Comments Using Logistic Regression." *Intelligence in Big Data Technologies—Beyond the Hype*. https://doi.org/10.1007/978-981-15-5285-4_20.

TABLES AND FIGURES

Table 1. BBC e-news dataset that contains the opinion of news that is published daily on the internet , that news mainly contains Business, Entertainment, Politice,Sports,Tech data on that opinion of the public in positive or negative .

NEWS CLASS	POSITIVE	NEGATIVE
Business	274	205
Entertainment	163	220

Politics	205	200
Sport	246	236
Tech	170	216

Table 2. Group statistical analysis of BERT with mean value of 83.20 and Linear Regression 77.80 and similarly the results of Standard Deviation and Standard Error Mean are given for NBERT and Linear Regression.

	Algorithm	**N**	**Mean**	**Std. Deviation**	**Std.Error Mean**
Accuracy	BERT	20	83.50	8.258	3.693
	LINEAR REGRESSION	20	77.80	8.167	3.652

Table 3. 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 Linear Regression with the value of $p > 0.001$)

		Lavene's test for equality of variances		**T-test for Equality of Means**					**95% confidence interval of the differenc**	
		F	**Sig**	**t**	**df**	**sig(2 tailed)**	**Mean diff**	**Std.error**	**Lower**	**Upper**
Accuracy	**Equal Variances assumed**	.053	.491	1.040	8	.039	5.400	5.194	-6.578	17.378
	Equal variances not assumed			1.040	7.99	.039	5.400	5.194	-6.578	17.378

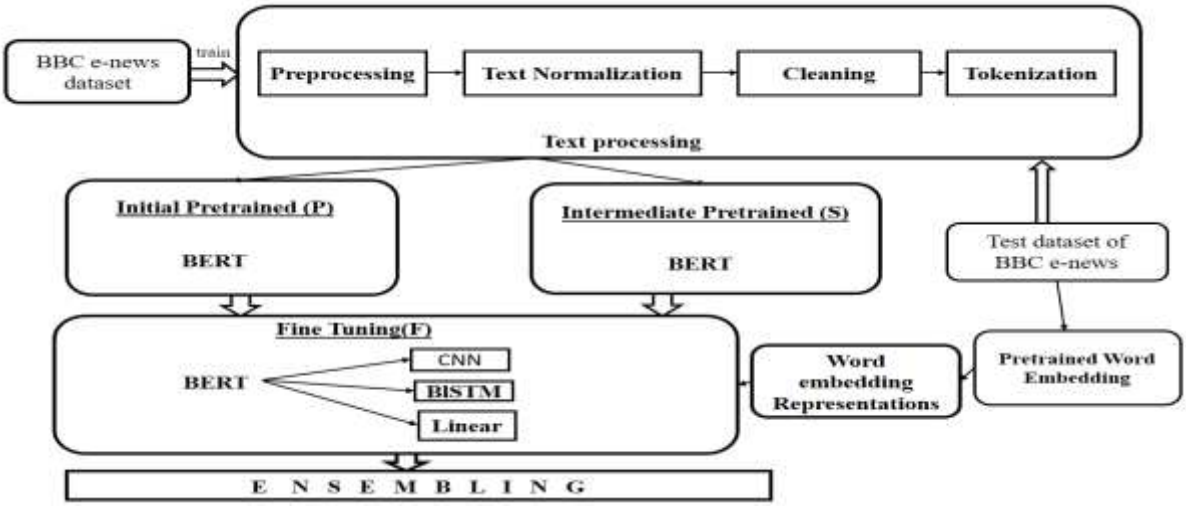

Fig. 1. Architecture for Sentiment analysis with Bidirectional Encoder Representation for Transformation on BBC-E news content reviews. This imports the text normalisation process to analyse the data and convert into word embedding to predict the analysis to produce the sentiment on given sentences in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

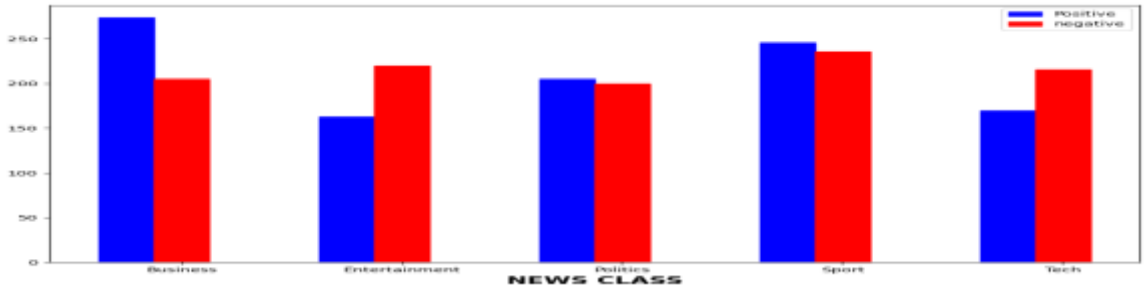

Fig. 2. Dataset that contains text forms with positive and negative opinion,that contains positive reviews and negative reviews on BBC e-news data. The main categories are Business, Entertainment, Politics, Sport, Tech.

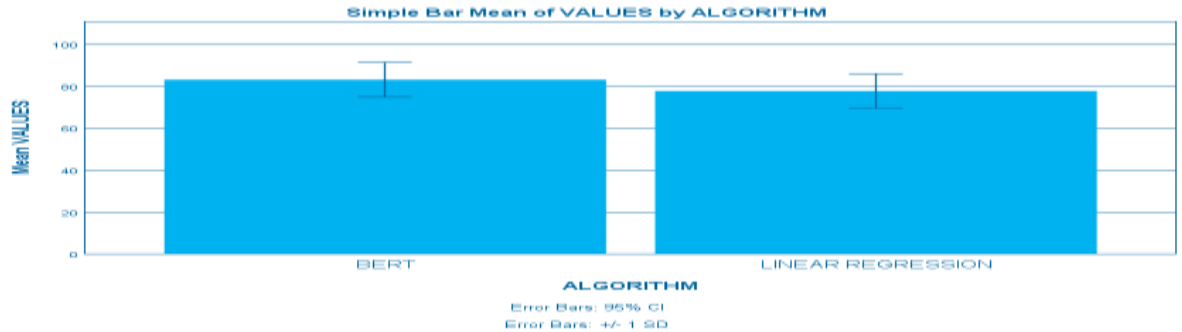

Fig. 3. Bar graph analysis of NovelBidirectional Encoder Representation for Transformation architecture and Linear Regression Algorithm architecture. Graphical representation shows the mean Accuracy of 83.5% and 77.80% for the proposed NBERT and Linear Regression Algorithm respectively. X-axis : BERT vs Naïve Bayes Classifier, Y-axis : Mean Accuracy ± 1 SD.