

Classify the Sentiments of email Contents using Novel Bidirectional Encoder Representation for Transformation over Naïve Bayes Algorithm

Chikkili. Hema Kumar¹, Malathi.K^{2*}

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, 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 Transformation (NBERT) over Naïve Bayes Algorithm.

Materials and Methods: Sample groups that are considered in the project can be classified into two, one for NBERT over Naïve Bayes Algorithm, which are tested using 0.80 for G-power to determine the sample size is 20 and for t-test analysis. Enron email content dataset that data collected from email.

Results: The automatic feature selection of NBERT algorithm splits the data with best fit, which has an average accuracy of 87%, which by far seems to be better than the Traditional Feature Extraction Method of Naïve Bayes Algorithm which gives 76%. The significance is around 0.0267 ($p < 0.05$) and therefore there is a statistically insignificant difference among the study group.

Conclusion: NovelNBERT seems to be better in finding the Sentiment in enron email dataset over the Naïve Bayes Algorithm.

Keywords: Novel BERT, Naïve Bayes Algorithm, Sentiment analysis, email, Spam Detection, Enron email dataset.

DOI: 10.47750/pnr.2022.13.S04.036

INTRODUCTION

As online social networks and communication is rapidly increasingly appealing to the public, extracting information and patterns from online documents and web content has become a burgeoning trend (Khan, Bashir, and Qamar 2014). Meanwhile, With the growing use of the Internet and digital devices, electronic mail (Email) has become a common means of communication and social networking. From 2011 to 2021, statistics indicated an increase of every year 45% in the number of global Email users. There are more than 4 billion email subscribers worldwide. By 2020, the number of email users had surpassed 4 billion, indicating that email was used by more than half of the world's population. The most recent estimate for 2021 is more than 4.1 billion people. In comparison to other media, electronic mail (Email) remains the most versatile social network platform. Statistics show that over 303.7 billion emails are sent every day in the business world (Farrell et al. 2021). Sentiment analysis, also known as emotional polarity computation and evolved from text mining, is a mining approach designed for determining the emotional state or attitude of writers or speakers toward specific topics or domains (Li and Wu 2010). Sentiment analysis for massive companies Emails may reveal useful business intelligence patterns. The methodologies and algorithms for sentiment analysis have been created and improved as the applications of sentiment analysis and opinion mining have been extended to numerous industries during the previous decade (B. Liu 2012, 2020). It is difficult to locate the correct information that the user need. We 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 an Email by categorising Email reviews (Joachims 2012).

In order to speed up procedures and lower the cost of developing and deploying containerized applications, many architectures and paradigms are being investigated. There are around 220 IEEE papers and 130 google scholar papers have been published over the past 5 years. The most cited article is "Sentiment Analysis for

Social Media using the Naïve Bayes Classifier of Machine Learning". In the existing research they didn't identify the availability and efficient prediction 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. Bidirectional Encoder Representations-BERT from Transformers is a state of art machine learning approach for natural language processing (Azzouza, Akli-Astouati, and Ibrahim 2020). BERT was introduced by Jacob Devlin and also 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 period to train and process successful Natural process models, even after preprocessing, full processing and tokenizing text reviews (Panchal 2019). The existence of a given feature in a class is assumed to be unrelated to the presence of any other feature by the Naive Bayes classifier. The Bayes rule, a means of looking at conditional probabilities that allows you to flip the condition around in a straightforward way, is at the heart of Bayesian classifiers (Troussas *et al.* 2013). The word negation refers to a statement that denies or reverses the value. 'Not' is the most common negation word (Kim 2018). The choice of a negative term in a movie review may have an impact since it alters the sentimental value. By using a negation word, positive words can become negative (Geetha and Karthika Renuka 2021). 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 for deploying sentences to analyse the data without loss, the system always uses a vectorization method. It is a probability technique, it presents the nearest value of expectation that leads to the data loss and accuracy of predicting the sentence of sentiment is slow. Also it checks the data in line form of head and tail. 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 Deep Learning Laboratory, Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS (Saveetha Institute of Medical and Technical Sciences). The proposed work contains two groups. Group 1 is taken as the novel Bidirectional Encoder Representation for Transformation and group 2 as Naïve Bayes Algorithm. The Novel BERT and Naïve Bayes Algorithm 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 Installation of the BERT model that can be used 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 Novel BERT model and Naïve Bayes Algorithm. 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 3.9.8.

Testing procedure for Sentiment Analysis with Novel BERT

NBERT-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.

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

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 analyzes the dataset into two parts in that train contains 28000 to train with the algorithm.

Step 4: Test Classifier

In the test classifier we 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 Algorithm

Input:- Enron email dataset, prediction statement.

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

```
Tokenizer← Novel 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

Naïve Bayes Classifier

Bayesian classifiers are based around the Bayes rule, a way of looking at conditional probabilities that allows to flip the condition around in a convenient way. A conditional probability is a probability that event X will occur, the evidence Y. that is normally written $p(X | Y)$. The Bayes rule allows to determine this probability when all have in opposite results and two components individually : $p(X | Y) = p(X) P(Y | X) / P(Y)$.

Process

Step 1. First train the system using a good and relevant corpus of mail.

Step 2. Now, take the testing corpus of email and take one email at time.

Step 3. Tokenize the email content and preprocess it.

Step 4. Taking all tokenized words, find the posterior probability to train the dataset.

Step 5. Now, find the sentiWordNet score of all tokenized words in particular e-mail and add its posterior probability.

Step 6. Compute the class conditional probability using Naïve Bayes Classifier.

Algorithm

Input:- Enron email dataset, prediction statement.

```
From sklearn.naive_bayes import multinomialNB
```

```
For i in dataset:
```

```
    Train,Test ← tf.keras.preprocessing() \\ it processing the data
```

```
End for
```

```
While pd in SVM:
```

```
    vectorizer ← sklearn.feature_extraction.text.SVMVectorizer()
```

```
    Train, test ← vectorizer.fit_transform()
```

```
clf ← sklearn.linear_model.LogisticRegression()
```

```
End while
```

```
From sklearn import metrics
```

```
Predicted =mnb.predict(x_test)
```

```
Accuracy_score = metrics.accuracy_score(predicted, y_test)
```

```
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
```

```
X_train, X_test , Y_train, Y_test =train_test_split(test_count, dataset)
```

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

```
    prediction_sentence
```

```
End for
```

Output:- sentence, Accuracy

End

Dataset is collected from Enron email Dataset. In the dataset 70% is used for training and 30% used for testing. The Novel BERT and Naïve Bayes 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 (Udhayakumar et al. 2021).

RESULTS

In this proposed system it was observed that Novel BERT uses optimization procedures to deliver high accuracy and no data loss with a bidirectional. Where Table 2 shows the data present in Enron email dataset that contain a positive and negative opinion information count, Fig. 2 represents a count of positive and negative opinion. Table 3 shows the statistical calculation such as mean, standard deviation and standard error mean for BERT and Naïve Bayes Classifier. The mean, standard deviation and standard error mean for BERT are 87.1, 1.8783, 1.840 respectively. The mean, standard deviation and standard error mean for Naïve Bayes Classifier are 76.4, 3.2738, 1.464 respectively. It is inferred that the mean accuracy for Novel BERT is 92 which is greater than the mean accuracy of comparison architecture which is 83. Moreover, the mean accuracy value of Bidirectional Encoder Representation for Transformation is 87.1 which seems to be superior to the Naïve Bayes Classifier. Figure 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 0.778 and 0.329, respectively and standard error difference and confidence interval are lower than Naïve Bayes Classifier. The mean accuracy is depicted in Fig. 3. For the supplied dataset of Sentiment Analysis on Enron email Dataset, BERT appears to be superior to movie review comments.

DISCUSSION

The proposed Bidirectional Encoder Representation for Transformation provides better containerized application deployment with analysis of sentences with word compared to Naïve Bayes Classifier Algorithm. The mean, standard deviation and standard error mean for Bidirectional Encoder Representation for Transformation deployment are 87.1, 1.8783, 1.840 respectively. The mean, standard deviation and standard error mean for Naïve Bayes Classifier deployment are 76.4, 3.2738, 1.464 respectively. It is inferred that the mean accuracy for Novel BERT is 87.14 which is greater than the mean accuracy of comparison architecture which is 76.4.

A systematic review on containerized application deployment techniques presents around 45 studies, the analysis of various papers shows that BERT is the most used algorithm for text classification (Rothman 2021). Since most of the movie reviews sentiment are related to text links (Ding, Qi, and Lin 2021), the algorithm provides the best accuracy for finding positive and negative comments. sentiment analysis with the BERT algorithm has more accuracy compared to the Naïve Bayes Classifier.

In industry sentiment analysis data, according to the results of the testing. The varied NBERT and Naïve Bayes Classifier performed the best among the model combinations created on all daous text representations and word embedding methods, the models that used the BERT word embedding method performed the best (Başarslan and Kayaalp 2021). Which have produced positive results in sentiment analysis, as well as the performance of transformers like RoBERT and DistilBERT in neural networks like LSTM and RNN (S. Liu, Tao, and Feng 2019). To train a sentiment classification model, we extracted tf-idf features from the Enron email dataset and fed them into three different machine learning classification algorithms. The sentiment classification findings for all tests were then compared (Chaudhary 2020).

The limitation of the study is scaled if the analysis is of various word forms and sentences, 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 BERT Bidirectional Encoder Representation for Transformers algorithm identifies the sentiment analysis of movies with better accuracy compared to the term Naïve Bayes Classifier algorithm. BERT algorithm that helps to function in both directions to predict the analysis on dataset with bidirectional encoder. And also BERT gives no loss of data than Naïve Bayes Classifier, it functions on the probability terms on data depends on X and Y value to analysis the sentiments.

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 KM 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, Melvinvíctor De Pours, 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. Azzouza, Nouredine, Karima Akli-Astouati, and Roliana Ibrahim. 2020. "TwitterBERT: Framework for Twitter Sentiment Analysis Based on Pre-Trained Language Model Representations." *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-3-030-33582-3_41.
5. Başarslan, Muhammet Sinan, and Fatih Kayaalp. 2021. "Sentiment Analysis on Social Media Reviews Datasets with Deep Learning Approach." *Sakarya University Journal of Computer and Information Sciences*. <https://doi.org/10.35377/saucis.04.01.833026>.
6. 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.
7. Chaudhary, Kartik. 2020. "Sentiment Analysis with Python: TFIDF Features." October 13, 2020. <https://dropsofai.com/sentiment-analysis-with-python-tfidf-features/>.
8. Deepanraj, B., N. Senthikumar, 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. Ding, Zhongxiang, Yali Qi, and Deping Lin. 2021. "Albert-Based Sentiment Analysis of Movie Review." *2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*. <https://doi.org/10.1109/aemcse51986.2021.00254>.

10. Farrell, Laura, Crystal R. Almond, Deborah S. Porterfield, Victoria Mobley, Sydney A. Jones, Marina Smelyanskaya, and Erika Samoff. 2021. "Automated Digital Notification of COVID-19 Diagnoses Through Text and Email Messaging - North Carolina, December 2020-January 2021." *MMWR. Morbidity and Mortality Weekly Report* 70 (46): 1603–7.
11. Geetha, M. P., and D. Karthika Renuka. 2021. "Improving the Performance of Aspect Based Sentiment Analysis Using Fine-Tuned Bert Base Uncased Model." *International Journal of Intelligent Networks*. <https://doi.org/10.1016/j.ijin.2021.06.005>.
12. 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.
13. Joachims, Thorsten. 2012. *Learning to Classify Text Using Support Vector Machines*. Springer Science & Business Media.
14. 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.
15. Khan, Farhan Hassan, Saba Bashir, and Usman Qamar. 2014. "TOM: Twitter Opinion Mining Framework Using Hybrid Classification Scheme." *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2013.09.004>.
16. Kim, Ricky. 2018. "Another Twitter Sentiment Analysis with Python — Part 5 (Tfidf Vectorizer, Model Comparison, Lexical Approach)." *Towards Data Science*. January 13, 2018. <https://towardsdatascience.com/another-twitter-sentiment-analysis-with-python-part-5-50b4e87d9bdd>
17. Li, Nan, and Desheng Dash Wu. 2010. "Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast." *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2009.09.003>.
18. Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.———. 2020. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
19. 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>.
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. 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.
28. Troussas, Christos, Maria Virvou, Kurt Junshean Espinosa, Kevin Llaguno, and Jaime Caro. 2013. "Sentiment Analysis of Facebook Statuses Using Naive Bayes Classifier for Language Learning." *IISA 2013*. <https://doi.org/10.1109/iisa.2013.6623713>.
29. 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. The data that was collected from Enron about email content in text normalisation with the public opinion about it in positive or negative way of comments. This dataset name is Enron email dataset.

S.No	Data_column	Label_coloum
0	Just finished making pancakes for breakfast.. Oh and the yummiest part, it comes with a free strawberry syrup!	Positive
1	I really don't like to share my hair but have to. :/ frustrated:(can't sleep	Negative
2	Freaking full of doubt.	Negative
3	Inspired by you!<5	Positive

4	11 days to go before marriage, need a leave to attend it	Positive
---	--	----------

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

Enron email dataset	Training	Testing
Positive	2900	1290
Negative	3590	1360

Table 3. Group statistical analysis of Bidirectional Encoder Representation with mean value of 87.14 and Naïve Bayes Classifier 76.4 and similarly the results of Standard Deviation and Standard Error Mean are given for NBERT and Naïve Bayes Classifier.

	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	BERT	20	87.14	1.8783	1.0840
	Naïve Bayes Classifier	20	76.46	3.2738	1.464

Table 4. Independent Sample T-test Results with confidence interval of 87.4% and level of significance greater than 0.05 (BERT appears to perform significantly better than Naïve Bayes Classifier with the value of $p > 0.001$)

	Equal Variances	Levene's Test for Equality of Variance		Levene's Test for Equality of Variance						
		F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper

Accuracy	Assumed	.046	.026	-1.32	8	.0460	10.500	1.3487	-6.7876	14.572
	Not Assumed			-1.32	6.376	.460	10.5040	1.34565	-6.6080	14.752

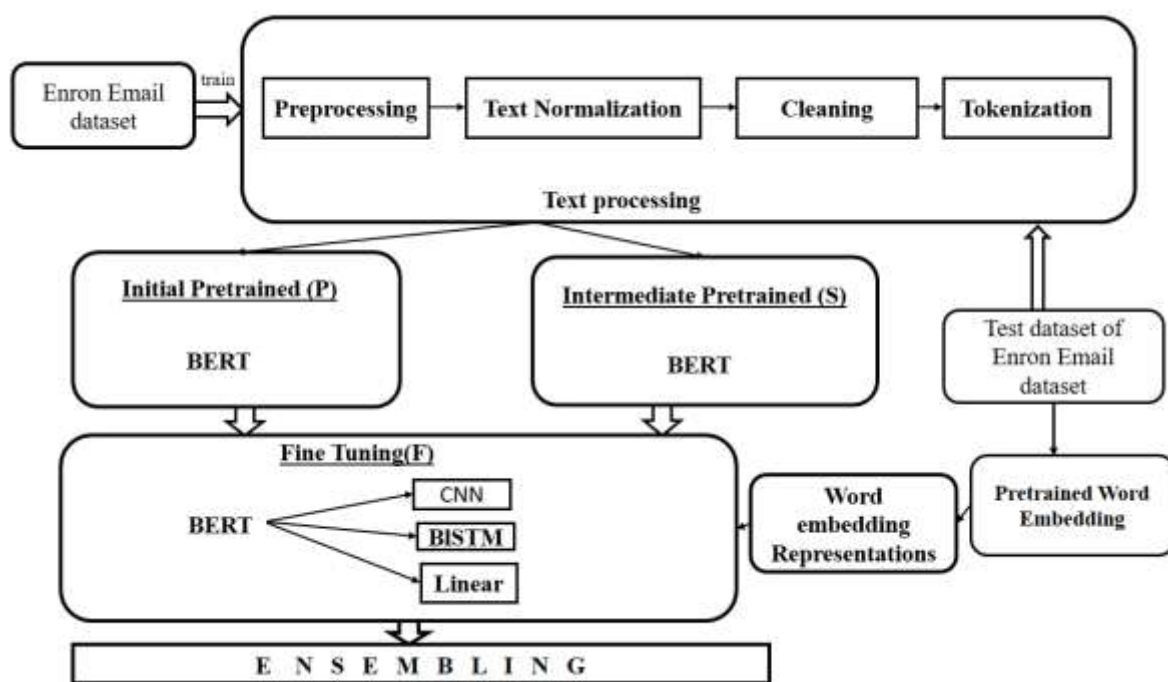


Fig. 1. Architecture for Sentiment analysis with Bidirectional Encoder Representation for Transformation on Enron email content reviews. This imports the text normalisation process to analyse the data and convert into word embedding to predict the analysis to produce the sentiment on given sentences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

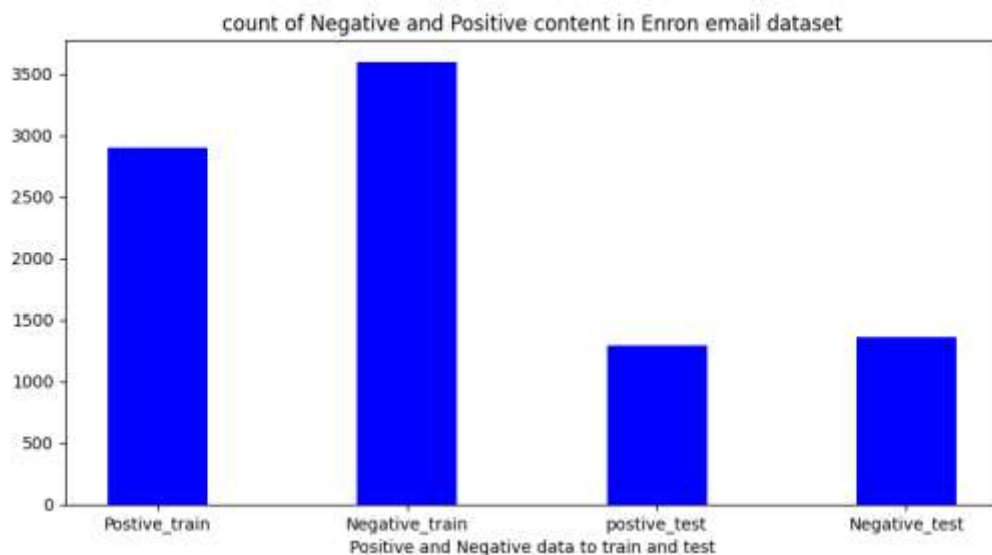


Fig. 2. Dataset that contains text forms with positive sentences on movie opinion with text representation and negative sentences on movie opinion with text representation. That contains positive content and negative content on Enron email dataset.

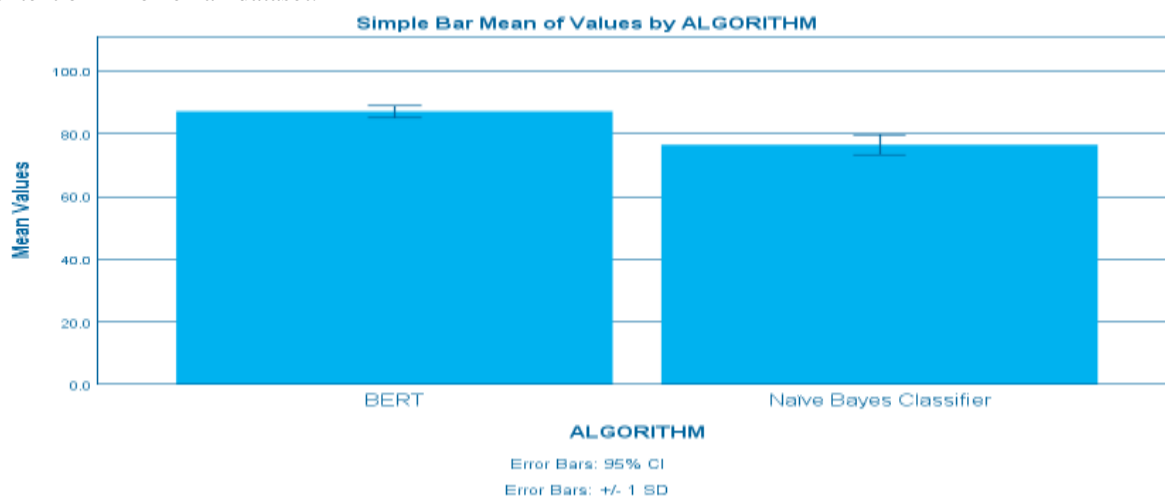


Fig. 3. Bar graph analysis of NovelBidirectional Encoder Representation for Transformation architecture and Naïve Bayes Classifier architecture. Graphical representation shows the mean Accuracy of 92% and 83% for the proposed NBERT and TF-IDF respectively. X-axis : BERT vs Naïve Bayes Classifier, Y-axis : Mean Accuracy ± 1 SD.