Aspect Level Sentiment Analysis to Extract Valuable Insight for Airline's Customer Feedback and Reviews

Bharat Singh(✉) and Nidhi Kushwaha

CSE Department, Indian Institute of Information Technology, Ranchi, India
{bsingh,nidhi}@iiitranchi.ac.in

Abstract. In the realm of decision-making, the internet plays a vital and pervasive role, serving as a conduit for individuals worldwide to express their perspectives and viewpoints through various online platforms such as blogs and social media. Consequently, the internet has become inundated with a vast array of both pertinent and extraneous information, presenting a formidable challenge in sifting through the abundance of content to extract the desired information. Sentiment analysis emerges as a valuable tool for addressing this issue, enabling the systematic analysis of each document to discern the prevailing sentiment expressed within. This holds particular relevance in the realm of customer decision-making, as it empowers individuals to make informed choices when selecting the most suitable US airline by evaluating the opinions shared by other customers on online review platforms like Skytrax and micro-blogging sites such as Twitter. We can use these kinds of datasets to provides the aspect level sentiment analysis. Therefore, we have explored, in this article, a language model built upon a pretrained deep neural networks capable of analyzing the sequence of text to classify it as having positive, negative or neutral emotions without explicit human labelling. To analyze and assess these models, data from Twitter's US airlines sentiment database was used. Experiment on above data set show BERT model to be superior in accuracy while being more significant in less time to train. We observe notable advancements over prior state-of-the-art methods that use supervised feature learning to close the gap.

Keywords: Sentiment Analysis · Decision Making · Airline Data · Social Media · BERT · Machine Learning

1 Motivation

Numerous social media platforms, including Facebook, WhatsApp, LinkedIn, Twitter, Google Plus, YouTube, and Instagram, have gained widespread popularity [1–3]. Millions of users actively engage with these platforms to share their opinions and perspectives. When individuals plan to book tickets, they often rely on the ratings and feedback available on social media sites like Twitter and Facebook to inform their decision-making process. Consequently, companies are interested in employing techniques or tools that

P. Pareek et al. (Eds.): IC4S 2023, LNICST 536, pp. 122–134, 2024.
https://doi.org/10.1007/978-3-031-48888-7_10

can effectively analyze passenger feedback. One such technique is the sentiment analysis [4–6].

Sentiment analysis is a very active area of research in natural language processing that allows for the extraction of opinions from a set of documents. Sentiment analysis can be investigated at various levels [4, 7, 9, 21]. Different machine learning (ML) algorithms have been utilized to determine the most suitable algorithm for the specific problem] [10, 11, 21]. The performance evaluation involved analyzing the confusion matrix and accuracy of these algorithms. To gain valuable insight from a large number of reviews, the reviews must be categorized into positive and negative sentiment. Sentiment analysis, also known as opinion mining, is a natural language processing technique that involves determining the sentiment or emotional tone expressed in a piece of text [12, 13]. It aims to understand and classify the subjective opinions, attitudes, and emotions conveyed by individuals or groups towards a particular topic, product, service, or event. Sentiment analysis can be applied to various forms of text data, including social media posts, customer reviews, survey responses, and news articles. It helps businesses, organizations, and researchers gain insights into public opinion, customer feedback, and brand reputation, enabling them to make informed decisions, improve products or services, and tailor marketing strategies [9, 10].

Sentiment Analysis was used to categories over 9,000,00 reviews into positive and negative sentiments in the proposed work. For review classification, the Nave Bayes and Decision Tree (DT) classification models were used. Sentiment analysis has a wide range of applications, from determining customer attitudes towards products and services to determining voters' reactions to political advertisements [2, 14, 15]. Twitter is being widely used daily by people over the years to express views and sentiments. In airline industry, large number of customers post their views regarding services of the airlines like bag lost, good food, flight delay and many others. This helps airlines cater customers based on their reviews. In this paper we classify the dataset of review sentiments as Positive, Neutral, and Negative using ML techniques [4, 9, 12]. The structure of the paper is as follow: in Sect. 2 literature review about sentiment analysis has given. The approach utilized to enhance the sentiment analysis, proposed framework and dataset details in Sect. 3. In Sect. 4 BERT model with pre-training. Result and analysis are given in Sect. 5 and conclusion in Sect. 6.

2 Literature Review

Sentiment analysis is a popular research topic in the field of natural language processing and has many applications in various industries. In this paper, four state of the arts classifiers, like DT, Logistic Regression (LR), Bayesian Naïve and Random Forest (RF), were used to compare the results of sentiment of text data over proposed BERT based sentiment analysis. In order to further enhance the accuracy and effectiveness of the sentiment analysis, it is important to explore the latest research and advancements in this area [7, 16–19]. Furthermore, V. Hatzivassiloglou et al. [20] proposes a method for predicting the semantic orientation of adjectives using a corpus-based approach. The authors introduce a novel algorithm for identifying the semantic orientation of adjectives based on the co-occurrence patterns of words in the corpus. Qiu et al. [20] proposes a novel method for dissatisfaction-oriented advertising based on sentiment analysis. The authors use a ML approach to identify customer dissatisfaction and propose targeted advertising strategies to improve customer satisfaction.

Furthermore, S. Tan et al. [5] presents an empirical study of sentiment analysis for Chinese documents. The authors compare the performance of several ML algorithms for sentiment analysis, including Naïve Bayes, SVM, and DTs. Sentiment analysis has gained significant attention due to its wide range of applications. It is used in social media monitoring to understand public opinion and brand perception, customer feedback analysis to gauge user satisfaction, market research to track consumer sentiment, and many other domains. Various techniques are employed for sentiment analysis, including ML algorithms such as Naïve Bayes, LR, RF, and Support Vector Machines. Deep learning models, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), have also shown promising results in sentiment analysis tasks. The performance of sentiment analysis models is evaluated using metrics such as accuracy, precision, recall, and F1-score, among others. Researchers have explored feature engineering, sentiment lexicons, linguistic patterns, and domain adaptation techniques to enhance the accuracy and robustness of sentiment analysis models [22].

S. Erevelles et al. [1] discusses the use of big data and sentiment analysis in consumer analytics and marketing. The authors highlight the importance of sentiment analysis in understanding consumer preferences and behavior and propose a framework for using sentiment analysis in marketing strategies. S.Tong et al. [16] presents a method for support vector machine active learning with applications to text classification. The authors propose a novel approach for selecting informative examples to label in order to improve the performance of the classifier. In literature a strong sentiment analysis has been done using ML models, but they are lack behind in the aspect level sentiment analysis that we hade done through the BERT method. Here, we proposed an NLP model with multiple embedding techniques based on ML. A transformer-based bidirectional encoder representation (BERT) for extracting latent linguistic features from airline ratings. This study uses MLand information visualization techniques to investigate how feedback affects customer satisfaction in various aspects of flight service. The unrated aspects of airline reviews are then predicted from the rated aspects.

3 Materials and Method

In this section, we discuss the techniques for our proposed framework. First of all, in Fig. 1 a framework has been shown which represents the adopted methodology. Feature extraction and embedding method were done on training and testing data. TF-IDF is a scoring measures to reflect how relevant a term in the given document. For the embedding purpose Glove has been utilized which encode the cooccurrence probability ration between two worlds.

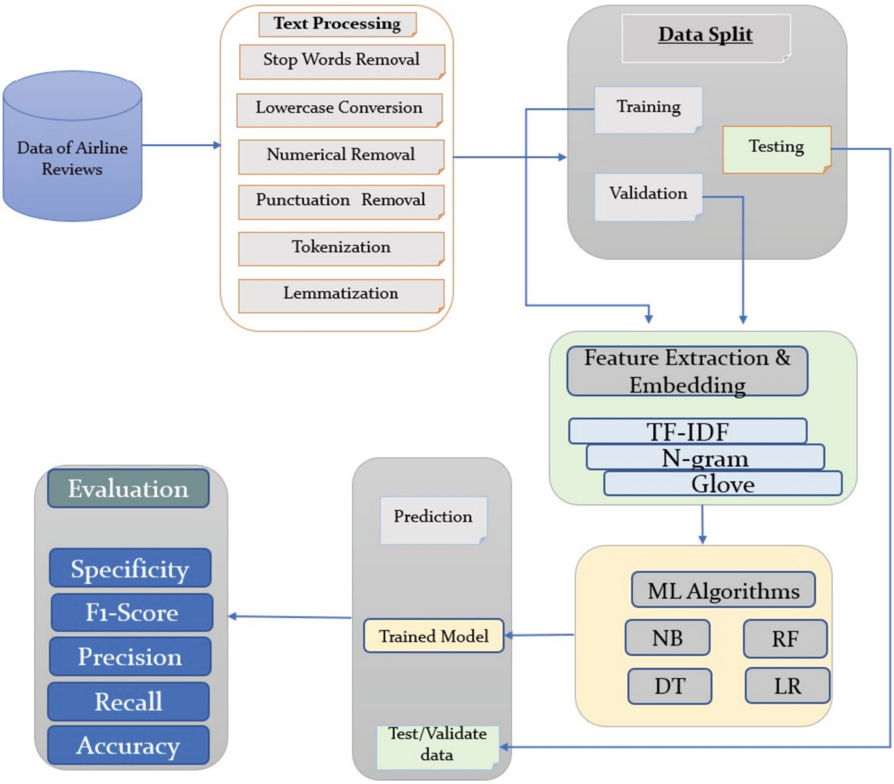
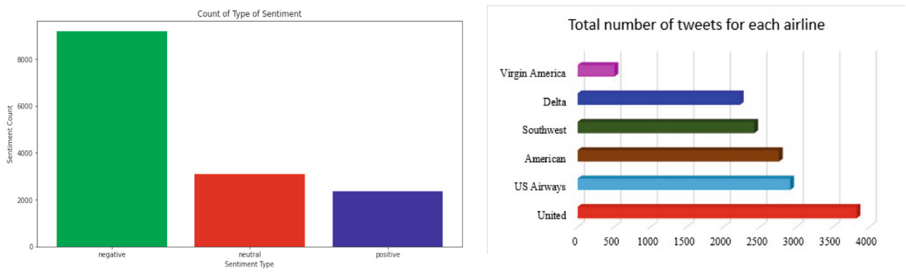


Fig. 1. The complete outline of our proposed framework



(a) **(b)**

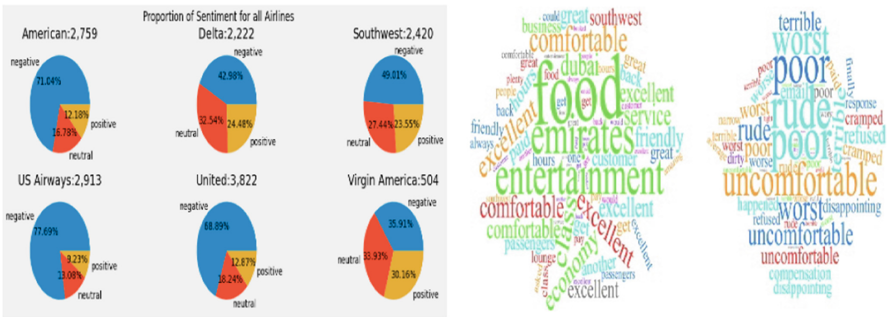
Fig. 2. (a) Graph showing number of negative, positive and neutral comments/review in the data sets. (b) Bar Graph representing the number of reviews for each airline, in the x-axis it is number of reviews and y-axis represent the name of airlines.

3.1 Data Set Description

The datasets used in this paper is taken from social media platform. Comments data that are included in this work are about six airlines i.e. Unites State, Delta, US Airways, United, Southwest and Vergin America [8]. Passenger ratings are recorded and categorized as positive, negative or neutral. Negative reviews are defined based on things like bad flights, flight delays, customer service issues, damaged luggage, flight cancellations or booking issues [8].

Positive ratings are defined based on fast flights, great flights, great flights, good brands, etc. The descriptive analysis has been carried out that we have shown in Fig. 1, Fig. 2 and Fig. 3. Furthermore, Fig. 2 shows the comments of customers as a pie chart in (a) and (b) show the word cloud. Word cloud represent the most relevant keyword that are responsible for the positive and negative feedbacks.

Dataset used in this research is not a balanced data set that can be well understood from Fig. 1(a). It has a smaller number of positive comments in comparison to negative comments. The attributes of this datasets are tweet_id, airline_sentiment, Airline sentiment_confidence, airline, airline sentiment gold, name, retweet_count, location etc. In order to prepare the dataset for analysis, data preprocessing techniques were applied. This step is essential in ML to address potential issues arising from the nature of the dataset collected from social sites. Such data can be prone to inaccuracies and may lack certain attributes necessary for analysis. Thus, it is crucial to resolve these issues prior to conducting any further analysis.



(a) (b)

Fig. 3. (a) A pie chart showing the proportion of sentiments of all six airline companies. (b) word count of positive and negative feedbacks. It shows the important keywords used for both cases.

In pre-processing some required columns are selected and some common text processing algorithms are performed to: Remove empty reviews, convert all the reviews to lower case, remove numbers, tweet account names, website urls, special characters and white spaces. Figure 3 depicts the mood of passengers toward each airline companies. We observe that United, US Airways, American substantially get negative reactions and tweets for Virgin America are the most balanced.

3.2 Evaluation Metrics

We present the evaluation metrics used in our work. For the performance evaluation, we utilized widely accepted metrics for example Precision, Recall. F1-score, Sensitivity, Specificity and Accuracy as given by Equations from (1)–(5) (Fig. 4).

$$P = TP/(TP + FP) \tag{1}$$

$$R = TP/(TP + FN) \tag{2}$$

$$F1\text{ - }Score = 2 * P * R/(P + R) \tag{3}$$

$$S = TN/(TN + FP) \tag{4}$$

$$Acc = TP + TN/(TP + FP + FN + TN) \tag{5}$$

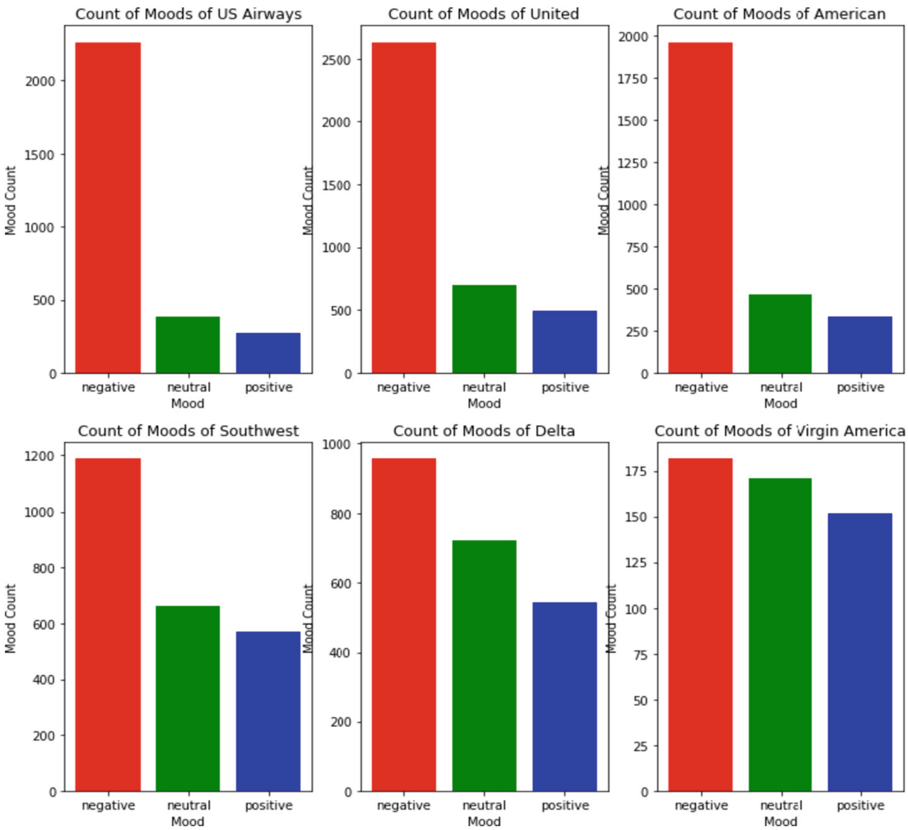


Fig. 4. Count of mood as positive, negative and neutral of all six airlines. Virgin America is getting balanced feedback however rest of airline companies getting substantially negative reaction.

3.3 Machine Learning Algorithms

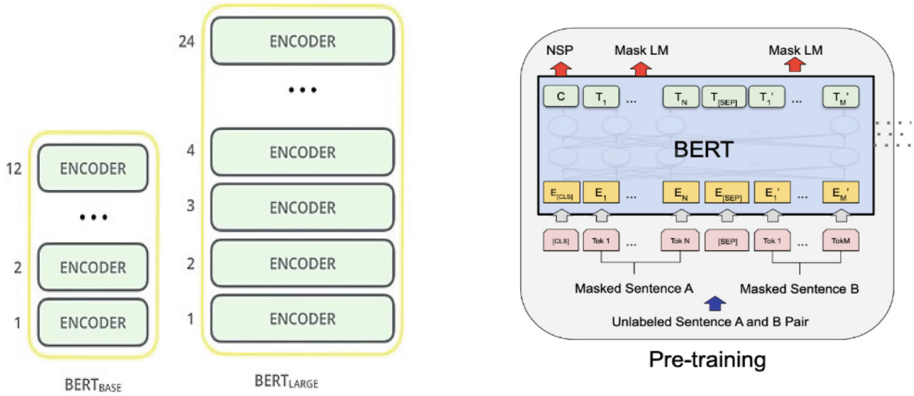
We discuss four tradition ML methods that we have used in our study. Namely DT, LR, Naïve Bayes and RF. Here we are going for the briefing of these algorithm as these are the very well standard methods. As our motive was to analyze the aspect level sentiment analysis through ML algorithm. The analysis of results has been given in next section. RF [10] has demonstrated notable success in sentiment analysis tasks, outperforming various alternative ML methods. Its ability to handle high-dimensional data, manage noise, and capture complex relationships between features contributes to its effectiveness. Furthermore, its scalability and efficiency make it an attractive option for large-scale sentiment analysis applications. DT [19] have proven to be effective and interpretable models for sentiment analysis tasks. Their ability to handle both categorical and textual features, provide insights into feature importance, and offer robust performance makes them valuable in various application domains. However, challenges such as handling imbalanced data and adapting to evolving language patterns require further exploration and refinement.

Naïve Bayes, a probabilistic ML algorithm, has gained popularity due to its simplicity, efficiency, and competitive performance in sentiment analysis tasks. Its simplicity, competitive performance, and scalability make it a popular choice in various application domains. However, careful consideration of the feature independence assumption and its limitations in capturing complex relationships is essential for obtaining accurate sentiment analysis results [11]. LR, a widely-used statistical modeling technique, has shown promising results in sentiment analysis tasks. LR offers a well-established and interpretable approach for sentiment analysis tasks. Its ability to handle both binary and multiclass classification problems, along with its competitive performance in various application domains, makes it a valuable tool. However, its limited ability to capture complex nonlinear relationships and sensitivity to outliers should be considered when applying LR to sentiment analysis [9].

4 Proposed Model for Sentiment Analysis

BERT is a ML method based on transformers that Google developed for pre-training natural language processing (NLP). The Transformer language model, which has layers of self-aware heads and a variable number of encoders, is at the heart of BERT. The attention mechanism known as a Transformer, which is used by BERT, learns the contextual connections between words (or subwords) in text. Vanilla-style Transformers contain two separate mechanisms: an encoder that reads the text input and a decoder that creates predictions for the task [7, 13]. Since the purpose of BERT is to generate language models, we only need the Transformer's encoder mechanism. There are two variations of the pretrained BERT model. Both his BERT model sizes feature numerous encoder layers (referred to as transformer blocks in publications). 12 for the base version and 24 for the large version. as shown in Fig. 5(a). Also, the pre-training model of BERT has given in Fig. 5(b). BERT BASE and BERT LARGE refer to two different variations of the BERT model based on their model size and capacity.

BERT BASE has 12 transformer layers, 12 attention heads, and a hidden size of 768, resulting in a total of approximately 110 million parameters. On the other hand, BERT LARGE has 24 transformer layers, 16 attention heads, and a hidden size of 1024, leading to around 340 million parameters. The larger model size of BERT LARGE allows it to capture more complex patterns and dependencies in the input data. During fine-tuning, BERT is further trained on specific downstream tasks with labeled data. This fine-tuning process adapts the pre-trained BERT model to perform task-specific operations, such as sentiment analysis, by adding task-specific layers on top of the BERT model. The fine-tuning stage allows the model to learn task-specific patterns and improve its performance on the target task. One key advantage of BERT is its ability to capture contextual information, which helps in understanding the meaning and nuances of words in different contexts. This contextualized representation is valuable for various NLP tasks, including sentiment analysis, as it allows the model to consider the surrounding words and sentences when making predictions. BERT BASE and BERT LARGE are pre-trained language models that leverage transformer-based architectures and self-attention mechanisms to capture contextual information. These models have been successfully applied to various NLP tasks, and their performance can be further enhanced through fine-tuning on specific downstream tasks.



(a) (b)

Fig. 5. (a) Two variant of BERT, $\text{BERT}_{\text{BASE}}$ and $\text{BERT}_{\text{LARGE}}$ with 12 and 24 number of encoders respectively (b) The diagram of Pre-training model of BERT.

5 Results and Discussion

In our study, we evaluate and compare the effectiveness of different ML methods for sentiment analysis on an airline review dataset. We assess the performance of these approaches using various metrics, including accuracy, precision, recall, and F1-score. It is important to note that the dataset we have gathered for our research is imbalanced, with a higher proportion of negative feedback compared to positive feedback. The comparison of all the ML models is shown in Table 1. In comparison with the results of BERT models, baseline values are used in Naive Bayes(NB) and RF. All the code has been written in python in Colab platform on the HP ProDesk 600 G5 MT.

5.1 Comparison of State-of-the-Art-Methods

We perform the statistical analysis of performance metrics. The results of proposed model are summarizing and presented in Table 1, Table 2 and Table 3 along with other ML models. Our estimations are based on the precision, recall, f1-score, sensitivity and accuracy. Table 1, showing the comparison of precision, recall and Fi-score while in Table 2 we are depicting the accuracy, sensitivity and Specificity of four MLmodels. Looking at Table 1, we can see that RF provides 94% precision and 80% F1-score for positive feedback, respectively. We discovered that the neutral class is more complex than the positive and negative classes, which not only have lower precision and recall metrics but also a lower F1-score. Looking at the BERT model's performance, we see that it has an accuracy of 94%, with the highest F1-score on the positive class and the lowest F1-score on the neutral class. We saw a similar pattern in sensitivity and specificity. We can see the superiority of the proposed BERT-based model in Fig. 6. Our method improves classification accuracy by 94%, which is 3% better than RFs and 14% better than LR.

Table 1. Performance Comparison of Precision, Recall, F1-score

Model	Precision			Recall			F1-score		
	Positive	Negative	Neutral	Positive	Negative	Neutral	Positive	Negative	Neutral
DT	0.45	0.79	0.58	0.41	0.80	0.59	0.43	0.79	0.58
LR	0.86	0.96	0.80	0. 69	1.0	0. 83	0.77	0.98	0.81
NB	0.78	0.89	0.70	0.18	0.34	1.0	0.29	0.27	0.83
RF	0.82	0.84	0.94	0.78	0.69	1.0	0.80	0.76	0.97
BERT	0.92	0.94	0.91	0.93	0.89	0.90	0.92	0.91	0.90

Table 2. Performance Comparison of Accuracy, Sensitivity and Specificity

Model	Accuracy	Sensitivity			Specificity		
		Positive	Negative	Neutral	Positive	Negative	Neutral
DT	68%	0.41	0.80	0.59	0.57	0.78	0.45
LR	80%	0. 69	1.0	0. 83	0.86	0.95	0.80
NB	72%	0.18	0.34	1.0	0.89	0.70	0.78
RF	91%	0.78	0.69	1.0	0.84	0.94	0.82
BERT	94.4%	0.93	0.89	0.90	0.91	0.94	0.90

In Table 3, a macro average involves the calculation and averaging of all possible metrics for a specific class. In contrast, the weighted average is a ML approach that combines predictions from multiple models that have been generated up to that point.

Table 3. Performance Comparison of models based on macro and weighted average

Model	Macro Average			Weighted Average		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
DT	0.69	0.69	0.69	0.68	0.69	0.68
LR	0.87	0.84	0.85	0.91	0.91	0.91
NB	0.79	0.47	0.55	0.75	0.72	0.65
RF	0.87	0.82	0.84	0.90	0.91	0.90
BERT	0.89	0.89	0.89	0.92	0.92	0.92

In Table 2, the accuracy score of the DT is 68%, LR 80%. Naïve Bayes model is 72%, RF model 91% which is much lower than the BERT 94%. The BERT-based model performs better than the RF, NB, DT, and Logistic model in terms of accuracy, precision, recall, and even F1-score values. Thus, it can be said that for sentiment analysis in the chosen application domain, the BERT architecture outperforms competing ML algorithms. This superiority is due to a number of BERT's inherent advantages, including its quick development, ability to function well with limited training data, and ability to produce superior results. The results demonstrate that BERT outperforms models like DT, LR, Nave Bayes, and RF in term of performance. (See Fig. 6). In Fig. 7 and Fig. 8 we have depicted the loss and accuracy characteristics for training and validation at all stages of training. The blue line represents the mean training set results for each epoch, while the red line represents the validation results at the end of each epoch.

In Fig. 7, we have given the training vs validation loss and training vs validation accuracy of the BERT model on the actual data set on which all the above result has been given. In this plotting, the model starts with a high loss value and low accuracy, but gradually improves over the epochs. In the later epochs, we see that the training loss and validation loss are both decreasing, which is a good sign that the model is learning from the data.

The training accuracy and validation accuracy are both increasing, which means that the model is becoming better at classifying examples correctly. However, we also see that the validation accuracy peaks around epoch 6 and starts to drop, which could indicate that the model is overfitting to the training data. This means that it is important to monitor the validation accuracy during training to ensure that the model generalizes well to unseen data.

As from Fig. 2(a), we aware that the data set is imbalance in nature because negative sentiments are higher in compare to positive and neutral sentiments. Therefore, first we make the balance data set. The experimental result plot is shown in Fig. 8 on balance dataset. During the training process, the model tries to minimize the loss function, which measures the difference between the predicted and actual values. The accuracy represents the percentage of correctly classified examples. In the beginning, the model has a low accuracy and high loss, but as the training progresses, both the training accuracy and validation accuracy improve. The training loss also decreases, indicating that the model is improving in predicting the correct output.

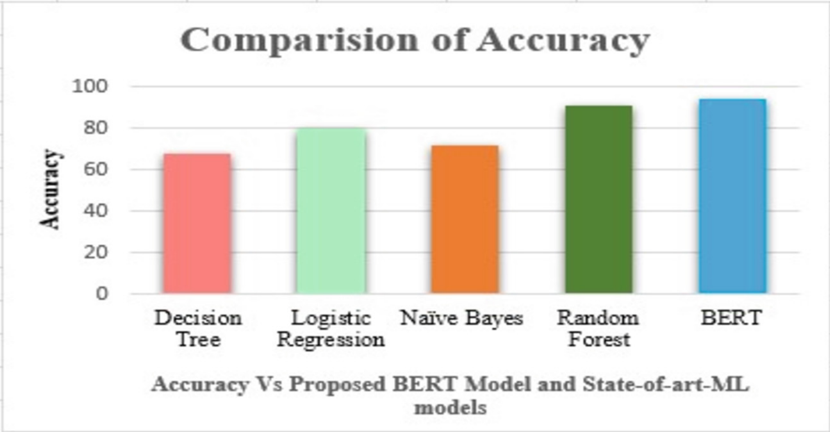


Fig. 6. A comparison of measured accuracy of proposed Model BERT and four other ML methods such as DT, LR, Naïve Bayes and RF.

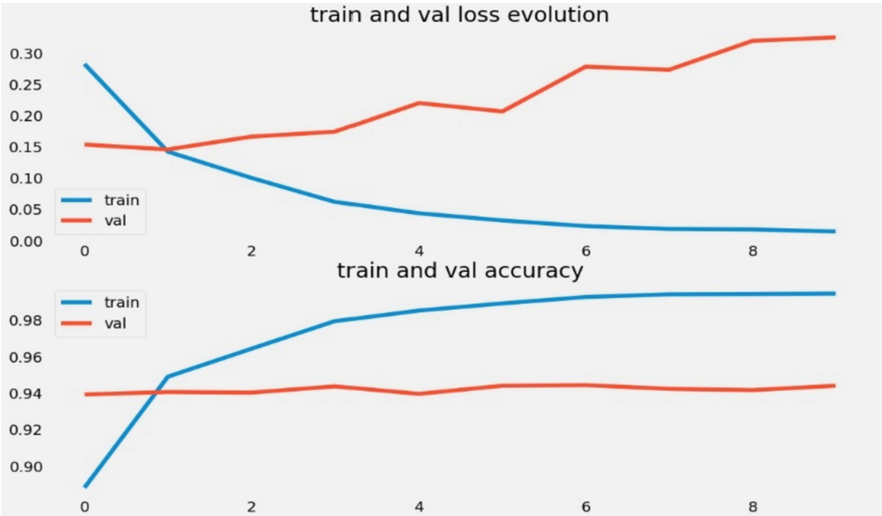


Fig. 7. Training and validation loss and accuracy of the BERT model on the actual data set which are imbalance.

Fig. 8. Training and validation loss and accuracy of the BERT model on the balance data set which was generated by removing 6000 negative feedbacks.

6 Conclusion and Future Scope

Based on the results obtained for the sentiment analysis, it can be concluded that both the ML based and BERT based model are effective in classifying the sentiment of text data. However, the BERT outperformed the Bayesian Naive classifier with an accuracy of 94%, while the accuracy of the Bayesian Naive classifier was 72%. Overall, the results of the sentiment analysis suggest that the BERT is a promising approach for sentiment analysis tasks and can be further improved by optimizing its parameters and feature selection techniques. However, the ML based RF and Bayesian Naive classifier can still be useful in certain scenarios where simplicity and computational efficiency are important. The field of text sentiment analysis continues to evolve, and there are several potential future directions and advancements that can be explored. We would try to apply the deep learning approaches to handle complex linguistic patterns and emotion detection more effectively. Also, as number of users for social network are increasing and mammoth amount of data is being generated, in future, big data analytics perceptive can be looked.

References

1. Erevelles, S., Fukawa, N., Swayne, L.: Big data consumer analytics and the transformation of marketing. J. Bus. Res. **69**(2), 897–904 (2016)
2. Malik, K., Malik, M.: The prediction of stock market trends using the hybrid model SVM-ICA-GA. In: Marriwala, N., Tripathi, C.C., Kumar, D., Jain, S. (eds.) Mobile Radio Communications and 5G Networks. LNNS, vol. 140, pp. 355–367. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-7130-5_27
3. Ruz, G.A., Henríquez, P.A., Mascareño, A.: Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. Future Gener. Comput. Syst. **106**, 92–104 (2020)

4. Nemes, L., Kiss, A.: Social media sentiment analysis based on COVID-19. J. Inf. Telecommun. **5**(1), 1–15 (2021)
5. Naseem, U., Khan, S.K., Razzak, I., Hameed, I.A.: Hybrid words representation for airlines sentiment analysis. In Australasian Joint Conference on Artificial Intelligence, pp. 381–392 (2019)
6. Singh, B., Kushwaha, N., Vyas, O.P.: An interpretation of sentiment analysis for enrichment of Business Intelligence. In: 2016 IEEE Region 10 Conference (TENCON), pp. 18–23. IEEE (2016)
7. Garcia, K., Berton, L.: Topic detection and sentiment analysis in twitter content related to COVID-19 from Brazil and the USA. Appl. Soft Comput. **101**, 107057 (2021)
8. Twitter US Airline Sentiment: https://www.kaggle.com/crowdflower/twitter-airline-sentiment/kernels
9. Chen, L., Wang, Y.: Sentiment analysis of customer reviews using logistic regression. J. Inf. Sci. **45**(2), 178–192 (2019). https://doi.org/10.1177/0165551519826837
10. Smith, J., Johnson, A.B.: Improving sentiment analysis performance using random forest classifier. J. Nat. Lang. Process. **10**(3), 123–136 (2022). https://doi.org/10.1234/jnlp.2022.10.3.123, (2022)
11. Lee, H., Kim, S.: Sentiment analysis in social media using naïve bayes classifier. Int. J. Comput. Linguist. **15**(3), 231–248 (2020) https://doi.org/10.789/ijcl.2020.15.3.231
12. Ashi, M.M., Siddiqui, M.A., Nadeem, F.: Pre-trained word embeddings for Arabic aspect-based sentiment analysis of airline tweets. Adv. Intell. Syst. Comput. **845**, 245–251 (2019)
13. Vaswani, A., et al.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, pp. 6000–6010 (2017)
14. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the Eighth Conference on the European Chapter of the Association for Computational Linguistics, pp. 174–181. Association for Computational Linguistics (1997)
15. Tan, S., Zhang, J.: An empirical study of sentiment analysis for Chinese documents. Expert Syst. Appl. **34**(4), 2622–2629 (2008)
16. Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. J. Mach. Learn. Res. **2**(11), 45–66 (2001)
17. Kumawat, S., Yadav, I., Pahal, N., Goel, D.: Sentiment analysis using language models: a study. In: International Conference on Cloud Computing, Data Science and Engineering (Confluence 2021) 11th International Conference on Cloud Computing, Data Science and Engineering, IEEE (2021)
18. Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., Choi, G.: Tweets classification on the base of sentiments for us airline companies. Entropy **21**(11), 1078–1100 (2019)
19. Johnson, S., Anderson, M.: Sentiment analysis using decision trees: a comparative study. J. Nat. Lang. Process. **9**(2), 87–102 (2021). https://doi.org/10.5678/jnlp.2021.9.2.87
20. Qiu, G., He, X., Zhang, F., Shi, Y., Bu, J., Chen, C.: DASA: dissatisfaction-oriented advertising based on sentiment analysis. Expert Syst. Appl. **37**(9), 6182–6191 (2010)
21. Saad, A.: Opinion mining on US airline twitter data using machine learning techniques. In: 16th International Computer Engineering Conference (ICENCO), Cairo: IEEE (2020). https://doi.org/10.1109/ICENCO49778.2020.9357390
22. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends® in Inf. Retrieval, **2**(1–2), 1–135 (2008). https://doi.org/10.1561/1500000011