An Enhanced Aspect-Based Sentiment Analysis Model Based on RoBERTa For Text Sentiment Analysis

Amit Chauhan[1], Aman Sharma[1], and Rajni Mohana[1,2*]
[1]Department of Computer Science & Engineering and Information Technology (CSE&IT), Jaypee University of Information Technology, Solan, HP 173234, India
[2]Department of Computer Science, Amity School of Engineering and Technology, Amity University Punjab, Mohali, Punjab 140306, India
E-mail: chauhanamit37@gmail.com, aman.sharma@juitsolan.in, rajni.mohanajuit@gmail.com*
*Corresponding author

Keywords: Sentiment analysis, NLP, BERT, ABSA, RoBERTa, XlNet

Received: November 15, 2023

Using an aspect-based sentiment analysis task, sentiment polarity towards specific aspect phrases within the same sentence or document is to be identified. The process of mechanically determining the underlying attitude or opinion indicated in the text is known as sentiment analysis. One of the most important aspects of natural language processing is sentiment analysis. The RoBERTa transformer model was pretrained in a self-supervised manner using a substantial corpus of English data. This means it was pretrained solely with raw texts and an algorithmic process to generate inputs and labels from those texts. No human labelling was involved, allowing it to utilise a vast amount of publicly available data. The authors of this work provide a thorough investigation of aspect-based sentiment analysis with RoBERTa. The RoBERTa model and its salient characteristics are outlined in this work, followed by an analysis of the model's optimisation by the authors for aspect-based sentiment analysis. The authors compare the RoBERTa model with other state-of-the-art models and evaluate its performance on multiple benchmark datasets. Our experimental results show that the RoBERTa model is effective for this important natural language processing task, outperforming competing models on sentiment analysis tasks. Based on the SemEval-2014 variant benchmarking datasets, the restaurant and laptop domains have the highest accuracy, scoring 92.35 % and 82.33 %, respectively.

Povzetek: Predlagan je izboljšan model analize sentimenta, ki temelji na aspektih (ABSA), ki uporablja RoBERTa in njene kontekstualizirane vgrajene predstavitve za izboljšano klasifikacijo sentimenta. Eksperimentalni rezultati kažejo večjo natančnost v primerjavi z najnaprednejšimi modeli, zlasti na nizih podatkov SemEval-2014, kar poudarja učinkovitost RoBERTa pri zaznavanju sentimentne polaritete specifične za posamezne aspekte.

1 Introduction

Natural Language Processing (NLP) is a subfield of artificial intelligence that focuses on the interaction between computers and human languages. NLP seeks to develop algorithms and models for human language analysis, comprehension, and production. NLP is used in various applications, including speech recognition, language translation, chatbots, sentiment analysis, and information retrieval. NLP combines machine learning, computer science, and language techniques to achieve these goals. Some of the main problems in NLP include the ambiguity and complexity of human language, handling grammatical and syntactic changes, and developing models that can capture the nuances of meaning and context in language. NLP has advanced recently despite these challenges, and it is expected to have a significant influence on the way writers use computers and technology. Beyond the advances in language creation and deep learning, there are many other exciting areas of NLP study. Among these is multilingual natural language processing (NLP), which aims to develop models and algorithms to understand and generate language in several languages. This has important implications for cross-cultural communication and global trade, where being able to understand and communicate in multiple languages is essential [10]. This is particularly important for sensitive applications such as healthcare, where it is essential to understand how a model arrived at a particular diagnosis or treatment recommendation. NLP is a fascinating field rapidly growing, with a wide range of potential uses and challenges to research. As this field grows, authors should expect to see more sophisticated and powerful language-based apps, which will transform the way authors interact with technology and communicate.

Sentiment analysis, also known as opinion mining, is a subfield of natural language processing that looks for and extracts subjective information from text. The field of sentiment analysis, as it is known to writers, began in the early 2000s, when academics began investigating the first machine learning algorithms to assess and classify sentiment in textual data. In 2002, Turney proposed the use of the supervised learning algorithm Naive Bayes for sentiment analysis, and this approach was widely implemented in the years that followed. In the mid-2000s, researchers began investigating the use of lexicons and sentiment dictionaries to improve the accuracy of sentiment analysis. attitude analysis is widely used in several areas, including marketing, customer service, and politics, to analyse public attitudes and opinions. As social media and online communication have risen in popularity, researchers are investigating new techniques, such as deep learning, to improve the accuracy and use of sentiment analysis in different scenarios [9].

A type of sentiment analysis called aspect-based sentiment analysis (ABSA) aims to ascertain people's opinions on specific attributes or parts of a product or service. Put another way, by going beyond straightforward emotion polarity classification, ABSA offers a more sophisticated understanding of the sentiment towards numerous components of a good or service [5]. This is important since reviews of products and services are often based on specific attributes, such as the quality of a smartphone's camera or the comfort of a car's seats. By performing sentiment analysis at the aspect level, businesses can gain a deeper understanding of the preferences and needs of their customers and make more informed decisions about marketing, customer service, and product development. As businesses strive to improve customer satisfaction and loyalty in a more competitive business climate, the significance of ABSA is growing. Numerous businesses, such as e-commerce, lodging, and healthcare, have used ABSA extensively. ABSA focuses on identifying the attitude towards specific attributes or characteristics of a product, service, or organisation [14, 13]. Customers often base their decisions on specific characteristics or features of a commodity or service, such as the cleanliness of a hotel room or the battery life of a smartphone, which makes this type of research essential. ABSA goes beyond simple sentiment polarity classification to give a more thorough understanding of the sentiment towards different components. Figure 1 describes the ABSA better with the help of an example. Standard procedures in ABSA include aspect extraction, sentiment polarity classification, and aspect-level sentiment aggregation. While aspect extraction involves identifying the aspects or features being discussed in the text, sentiment polarity classification establishes whether a certain aspect is being perceived positively, negatively, or neutrally. component-level sentiment aggregation combines the sentiment polarity ratings for each component to get an overall sentiment score for the good, service, or institution [17, 8].

Applications for ABSA can be found in many industries, including e-commerce, lodging, and medical. The diversity and richness of human language make ABSA a challenging task in natural language processing. One of the main problems is aspect extraction, which involves identifying the aspects or elements discussed in the text. This can be challenging since different people may refer to the same thing by different names, and different contexts may lead to different meanings for the same term. Another challenge is sentiment polarity classification, which requires understanding the nuances of language and context to determine the sentiment towards each component accurately. In terms of [25, 18] general emotion polarity, the statement "the battery life is okay" could be categorised as neutral. Still, if the battery life was previously perceived as inadequate, it could be positively polarised. Despite these challenges, recent advancements in natural language processing and machine learning have greatly improved the precision and effectiveness of ABSA. The availability of pre-trained models and large-scale annotated datasets is expanding, facilitating businesses' adoption of ABSA in their operations.

1.1 Contribution

In this paper, the authors present a novel method for aspect-based sentiment analysis using RoBERTa (ABSA-RoBERTa). Our approach is motivated because aspect and sentiment phrases in opinion articles often occur together and have positional interdependence. Furthermore, consumers can characterise traits in various ways, establishing semantic relationships. The authors argue that, unlike previous research that requires complex fine-tuning procedures for RoBERTa to account for these features, the ABSA-RoBERTa method naturally integrates these dependencies. Consequently, our model requires minimal fine-tuning for the next assignment to yield state-of-the-art results on benchmark datasets. Therefore, combining our approach with RoBERTa points to a promising new direction for aspect-based sentiment analysis.

1.2 Structure of paper

In section one, the introduction is described, starting with the NLP introduction, followed by the contribution and motivation of the study. In section two, related work is presented, along with a literature survey where extensive literature is discussed. Section three covers the proposed approach, detailing the dataset, the job done, and the preliminaries. Section four outlines the evaluation metrics used in the study. In section five, the authors discuss the results and provide insights regarding the proposed outcomes. Section six features the conclusion and discusses the future scope outlined by the authors.

2 Related work

Heng Yang et al.[23] 2021 show that the implicit aspect sentiments are typically dependent on the sentiments of the surrounding aspects, meaning that they can be recovered through aggregation, a type of dependency modelling. To

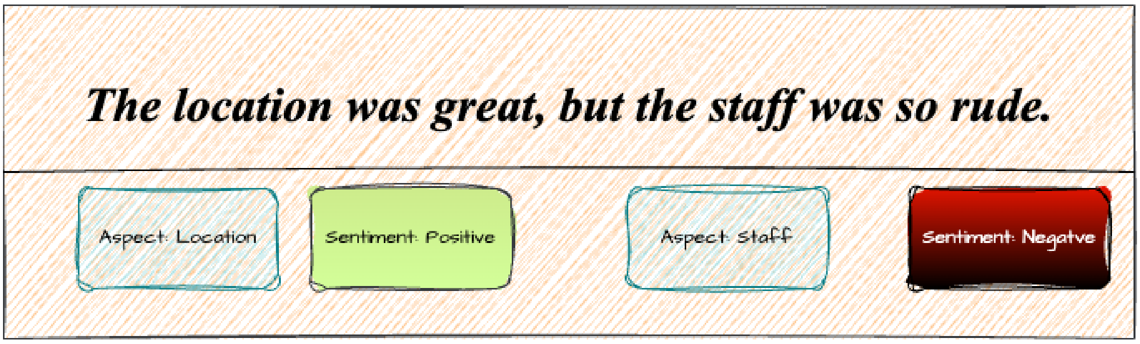

Figure 1: Example of ABSA

validate their findings on the SemEval 14 dataset, they employed the LSA+DeBERTa-V3-Large model. Heng Yang et al. [24] 2019 claimed that the two stages of natural language processing (NLP) are polarity categorization and aspect extraction. The LCF-ATEPC model can operate synchronously on both the Chinese Review dataset and the SemEval 14 dataset. Emanuel H. Silva et al. [19], 2021 indicated that the BERT-based models perform well in tasks requiring a profound comprehension of language, such as sentiment analysis. They developed a new approach for downstream tasks by adopting a Decoding-enhanced BERT with Disentangled Attention DeBERTa model using the SemEval 14 dataset to improve this theory further. Yiming Zang et al. [27], 2022 The authors claim that the absence of annotated data significantly impedes the creation of ASBA tasks. They developed a Dual-granularity Pseudo Labelling (DPL) to address this job. DPL provides a general framework that can be used to combine previous approaches from the literature for the same dataset, SemEval 14. Junqi Dai et al. [4] examined the dependency parsing trees over several well-known ABSA models and the inductive trees from the Pre-trained models (PTM). The authors found that, when tested on six SemEval datasets (14, 15, and 16), the inductive tree of fine-tuned RoBERTa performed the best and was more sentiment-word-oriented. Boauthorsn Xing et al. [22]; 2021 created a novel Aspect-level Sentiment classification model (ASC) with the following features: a dual syntax graph network that combines both types of syntactic information to comprehensively captures sufficient syntactic information, a knowledge integrating gate that re-enhances the final representation with further needed aspect knowledge; and an aspect-to-context attention mechanism that aggregates the aspect-related semantics from all hidden states inside the final representation. Alexander Rietzler et al. [16] 2020; Identified a model called Aspect-Target Sentiment Classification (ATSC) that suggests cross-domain Bert models outperform robust baseline models like Bert base models. Akbar Karimi et al. [11] 2021; It has been suggested that aspect extraction and aspect-target sentiment classification tasks can be handled with Parallel Aggregation and Hierarchical Aggregation without requiring fine-tuning the Bert base models in the SemEval 14 dataset. Youauthorsi Song et al. [21] 2019 demonstrated that using a pre-trained Bert model on the SemEval 14 dataset, which was a lighter model than the other models mentioned in this literature, the Attentional Encoder Network (AEN) performs better than the Recurrent Neural Network (RNN).

3 Proposed approach

This research presents a novel method for aspect-based sentiment analysis utilising the RoBERTa. Figure 2 displays the flowchart of the whole suggested model. Figure 2 outlines our methodology for predicting sentiments. Preprocessing the data was the initial step, and it was upon this that the most critical phase—aspect extraction—was completed. Authors can accurately anticipate a text's sentiment once the model has identified its constituent parts.

3.1 Dataset

SemEval 14 Task 4 Subtask 2 [12] is the data set used in the suggested method. SemEval, an annual international symposium on semantic evaluation, aims to evaluate NLP systems' efficacy. The purpose of Task 4 Subtask 2 in SemEval 2014 was to classify tweets' emotions better accurately. Participants were asked to classify the attitude expressed in tweets into one of five categories: "positive," "neutral," and "negative." The task was challenging due to the informal nature of tweets, which typically contain grammar errors and other noise. Many machine learning methods, including support vector machines and deep neural networks, were applied to complete this subtask. The results showed that natural language processing systems are still challenging. As we know, the ABSA is a challenging task on its own. The SemEval evaluation has been instrumental for researchers focusing on Aspect-Based Sentiment Analysis (ABSA), as high-quality datasets are crucial for such tasks. The data set is categorized into two domains: restaurant and laptop reviews, each further divided into positive, negative, and neutral sentiment classes.

1. For restaurant reviews: Positive: 728 training samples and 2,164 test samples
   Negative: 867 training samples and 196 test samples
   Neutral: 637 training samples and 196 test samples

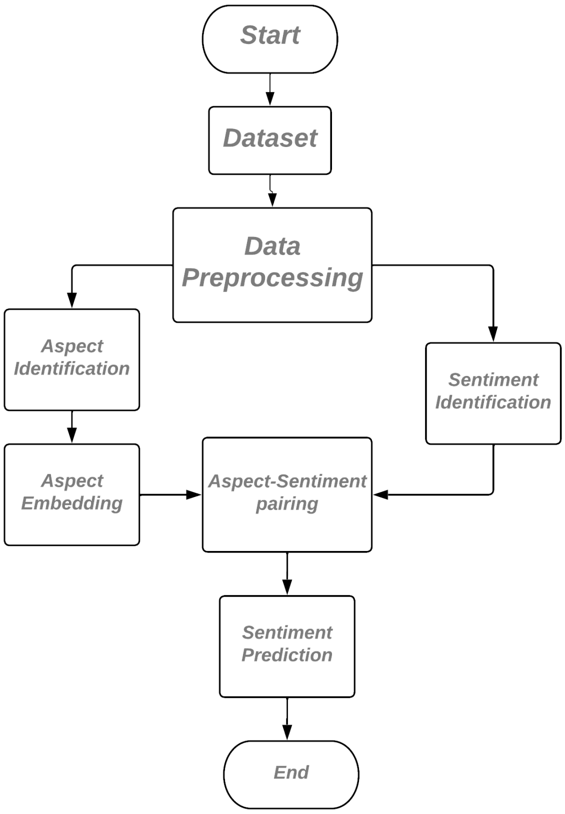

Figure 2: Proposed methodology for ABSA using RoBERTa model

2. For Laptop reviews: Positive: 341 training samples and 994 test samples
   Negative: 870 training samples and 128 test samples
   Neutral: 464 training samples and 169 test samples

This data set serves as a valuable reference for evaluating models in ABSA tasks.

3.2 Embedding layer

Embedding layers are fundamental to many natural language processing (NLP) models. These layers represent words or phrases as vectors in a high-dimensional space, and the interactions between the vectors capture the semantic meaning of the words or phrases. Training in massive corpora of text data using techniques like word2vec or GloVe is a common step in learning embedding layers. After that, the generated embedding can be used as input for other downstream NLP tasks like sentiment analysis or language translation. During training on a specific task, the embedding layers can also be better modified to capture the distinct nuances of the text data.

3.2.1 Glove embedding layer

In natural language processing, the glove [15] is an unsupervised learning technique that creates vector representations of words. These vector representations, or embeddings, capture the semantic meaning and context of words within a specific corpus. GloVe is often used to collect client attitudes and views regarding particular features or elements of a good or service in sentiment analysis activities such as Aspect-Based Sentiment Analysis (ABSA). Conversely, XLNet is a state-of-the-art language model that pre-trains using an auto-regressive technique on large amounts of text data. XLNet has been shown to outperform previous language models, such as BERT and GPT-2, in several natural language processing tasks, including ABSA. By integrating GloVe embedding with XLNet for ABSA, customer sentiment towards particular product or service elements may be further precisely and nuancedly evaluated. Combining the benefits of both algorithms allows ABSA models to more fully understand the relationships between words and the emotions they evoke, yielding more insightful and practical results for businesses.

3.3 RoBERTa

RoBERTa is a reimplementation of BERT that includes a setup for RoBERTa pre-trained models and minor adjustments to the significant hyperparameters and embedding. We don't need to utilise token type IDs or specify which token belongs to which segment in RoBERTa. The segments are readily divided with the help of the tokenizer.sep token (or) separation token.

3.4 Preliminaries

The SVM [7]puts the data in a high-dimensional space, and the model creates support vectors that help forecast the target labels by drawing a straight line, known as a hyperplane, to split the data into many classes. Equation (1) expresses the SVM classifier, and Equation (2) expresses the SVM classification for dual creation.

$$min_{f\zeta_i} \|f\|^2{}_k + C\sum_{i}^{l} \xi_i y_i f(x_i) \geq 1-\xi_i,\ for\,all\ i\xi_i \geq o \tag{1}$$

$$min_\alpha \sum_i \alpha_i - \frac{1}{2}\sum^{1}\sum^{1} \alpha_i\alpha_j \dagger_\rangle \dagger_| K(x_i,x_j) 0 \leq \alpha_i \leq c,$$
$$: for\,all\ i; \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{2}$$

The error generated at position (x_i, y_i) is measured by slack variables i in Eqs. (1) and (2), where i is the Langlier's multiplier. Random Forest (RF) [1] constructs a decision tree for every training set, averages those decision trees, and lets users select their preferred prediction outcome to anticipate the target labels. The RF classifier is provided by Equation (3).

$$\begin{aligned} \underline{r}(X) &= E_\theta\left[r_n(X,\theta)\right] \\ &= E_\theta\left[\frac{\sum_{i=1}^{n} Y_i 1_{[X_i\in A_n(X,\theta)]}}{\sum_{i=1}^{n} 1^* 1_{[X_i\in A_n(X,\theta)]}} 1_{E_n(X,\theta)}\right] \end{aligned} \tag{3}$$

In Eq. (3), $r_n(X,\theta)$ is the randomised tree of the rectangular cell of the random partition containing $E_n(X,\theta)$ trees. Long Short-Term Memory, often known as LSTM [26], is a type of recurrent neural network (RNN) design intended to manage long-term dependencies and avoid the vanishing gradient issue that certain traditional RNNs may experience. In LSTMs, a memory cell—a part that stores information over time—is employed with input, forget, and output gates. The LSTM can store and retrieve data selectively as needed thanks to these gates, which control the flow of information into and out of the memory cell. There are three examples of vanilla LSTM classifiers: (4), (5), and (6).

$$\delta W_* = \sum_{t=0}^{T}\left(\delta\star^t, x^t\right) \tag{4}$$

$$\delta R_* = \sum_{t=0}^{T-1}\left(\star^{t+1}, y^t\right) \tag{5}$$

$$\delta b_* = \sum_{t=0}^{T}\delta p_0 \tag{6}$$

where b is the bias weight, p is the peephole weight, R is the recurring weight, and W is the input weight. The BERT (Bidirectional Encoder Representations from Transformers) is used to learn and represent the contextualised meaning of words in a phrase [20] model with prior training. BERT-SPC can accurately classify the sentiment polarity (positive, negative, or neutral) of a given text input. It has been shown to outperform traditional machine learning models in several benchmark datasets used for sentiment analysis. BERT-SPC is frequently employed in social media monitoring, customer feedback analysis, and market research. The BERT objective function is provided by Equations (7) and (8).

$$L(\theta) = -\sum_{i=1}^{c} y^c \log\left(y^c\right) + L_{lsr} + \lambda\sum_{\theta\in\Theta}\theta^2 \tag{7}$$

$$L_{lsr} = -D_{kl}\left(u(k)\|p_\theta\right) \tag{8}$$

Where $\left\{y^{\text{authors}} \in \mathrm{R}^c\right\}$ is the output layer's anticipated sentiment distribution vector, y is the ground truth represented as a one-hot vector, λ is the coefficient for the L2 regularisation term, is the parameter set, and p is the network's predicted distribution. The DeBERT [6] allows the model to focus on different input aspects independently by using disentangled attention. To achieve this, the attention mechanism is split into multiple heads, each focusing on a distinct subset of the input. By doing this, DeBERTa gains enhanced capability to handle long-range dependencies and detect more subtle word connections. The model's decoder module explicitly uses the self-attention process to provide each word in the input sequence with a contextualised representation. Eq. provides the DeBERTa classifier (9).

$$A_{i,j} = \left\{H_{i'}P_{i|j}\right\}^* \left\{H_{j'}P_{j|i}\right\}^T \tag{9}$$

where H_i denotes the content vector of token i, and Pi|j denotes the relative position vector between tokens i and j. In this example, i and j represent two tokens in a phrase. XLNet randomly permutes the input sequence, and the model is trained to predict the probability of each permutation. This allows XLNet to record more complex word interactions and better manage long-distance dependencies. Another key component of XLNet is using a segment-level recurrence mechanism, which allows the model to consider previous input segments while forecasting the next word. Consequently, the model exhibits superior performance across several natural language processing tasks and has a more remarkable ability to represent long-term dependencies. In Eq. (10) is the XLNet objective function.

$$\max_0 E_z \sim Z_T\left[\sum_{t=1}^{T}\log p_\theta\left(x_{z_t} \mid x_{x<t}\right)\right] \tag{10}$$

Where x is the text sequence for which p_θ is the likelihood factorization order for an order z at a time t.

4 Evaluation metrics

The authors of this article employed the F1 Measure and accuracy to measure and compare the outcomes, as shown in the assessment tables and graphs of our study that are presented ahead of time.

1. Accuracy- Accuracy can be calculated as given in Eq. (11)

$$A = \frac{T_p + T_N}{T_P + T_N + F_P + F_N} \tag{11}$$

A fundamental indicator called accuracy measures the proportion of instances that were correctly predicted, where TP and TN represent the number of instances that correctly predict a label to be positive or negative, respectively. FP and FN represent the number of cases with incorrectly predicted labels. However, as researchers are usually more interested in the minority class than the majority class, accuracy is not the best choice for data sets that have imbalances. Great accuracy may sometimes reflect the accuracy of the majority class or both classes taken together rather than always indicating high accuracy for the minority class.

1. F1 Score, Accuracy, and Recall Precision quantifies the percentage of labels that the model correctly anticipated. The recall is the proportion of all pertinent labels the model identified correctly. As shown in equations (12), (13) and (14), precision and recall can be computed.

$$\beta = \frac{T_p}{T_P + F_P} \tag{12}$$

$$\gamma = \frac{T_p}{T_P + F_N} \tag{13}$$

Using the results from β and γ , F1 is given as in Eq. (14)

$$F1 = \frac{2 \times \beta \times \gamma}{\beta + \gamma} \tag{14}$$

5 Results and discussions

A method in natural language processing called aspect-based sentiment analysis (ABSA) determines how customers feel about particular features of a good or service. It can track client opinions on social media, increase customer happiness and loyalty, and help businesses make better decisions. The authors changed the data format for the class sentiment, which comprised three attributes: Positive, Negative, and Neutral, after tokenizing the provided text data using the TF-IDF and glove embedding techniques during the data-preprocessing phase.

A crucial step in aspect-based sentiment analysis (ABSA), which aims to pinpoint the exact characteristics or features of a good or service under evaluation, is aspect identification. The authors employed rule-based and machine learning-based approaches to identify specific elements of our methodology precisely. Authors originally developed a set of guidelines based on part-of-speech tags and grammatical dependencies to identify aspects. The authors then used a machine learning-based approach based on RoBERTa to improve aspect recognition accuracy further. The authors improved the ABSA Task with the RoBERTa model to identify attributes on a dataset of product evaluations, and they achieved state-of-the-art performance on several benchmark datasets.

Aspect-based sentiment analysis (ABSA) represents features or attributes of an item or service in a low-dimensional space by utilising the idea of aspect embedding. Utilising TF-IDF and GloVe word embedding, our approach to aspect embedding involved using RoBERTa, a pre-trained language model, to construct contextualised representations of texts' aspects. The authors refined the RoBERTa model to acquire aspect embedding, which they subsequently used to classify the sentiment associated with each aspect on a dataset of restaurant reviews [2, 3]. Our approach achieved state-of-the-art performance for aspect embedding in ABSA, outperforming earlier approaches already used on several benchmark datasets. Precise aspect embedding is essential for successful ABSA since it allows us to pick up on the subtleties of client mood.

A crucial aspect-based sentiment analysis (ABSA) step, which involves determining the sentiment expressed towards a particular element or characteristic of a good or service, is aspect sentiment pairing. Nouns in ABSA often denote aspects, whereas adjectives or adverbs denote moods. Finding the adjectives or adverbs in a sentence that best describes a particular aspect or feature is known as aspect sentiment pairing. This job is difficult since adjectives and adverbs can describe various traits or aspects, and the attitude towards a particular feature can change depending on the situation. Accurate aspect sentiment pairing is necessary to produce fine-grained sentiment analysis and give businesses insights into the specific features of their goods or services.

5.1 Model comparison

The authors performed training and testing on the SemEval 14 dataset for 6 different models, whose comparison is shown in Table 1. The observation indicates that the RoBERTa model outperformed all other approaches in the Restaurant and Laptop training datasets. It was also observed that the other models, which were not part of deep learning models, performed poorly in identifying the sentiment with the help of aspects. However, they showed an improved performance when the prediction of the feelings was carried out without recognising the aspects. The main comparison is between the two similar state-of-the-art models, BERT-SPC and RoBERTa. RoBERTa outperformed BERT-SPC by giving a 7.89 % increase in accuracy in the case of the restaurant dataset.

Table 1: Comparison of evaluation metrics of various models on the Restaurant dataset

Model	Accuracy (%)	F1 Score (%)
Random Forest	84.67	79.56
SVM	84.18	79.23
Naive Bayes	83.97	77.50
LSTM	70.88	67.45
BERT-SPC	84.46	76.98
RoBERTa (Proposed)	**92.35**	**91.05**

Table 2 shows the comparison of our model with other state-of-the-art models. The authors compare the excellent models with some currently top-performing state-of-the-art models for which our model outperformed the others. Our proposed model improved by 2.02 and 3.88 for the restaurant and laptop, respectively, over the current best model LSA+DeBERTa-V3-Large for the Restaurant dataset for SemEval 2014. The performance of different machine learning models on a given task can be assessed using a comparison table of accuracy and F1 scores for multiple

Table 2: Comparison of proposed model with baseline models

Model	Restaurant		Laptop	
	Accuracy (%)	F1 Score (%)	Accuracy (%)	F1 Score (%)
De-BERTa [8]	89.46	-	82.76	79.36
InstructABSA [5]	89.76	92.76	88.37	92.30
LSA+DeBERTa-V3-Large[6]	90.33	85.78	86.21	83.87
LCF-ATEPC [7]	90.18	85.88	82.29	85.29
BERT-SPC [14]	84.46	76.98	78.99	75.03
RoBERTa (Proposed)	**92.35**	**91.05**	**82.33**	**84.04**

models. The proportion of accurate predictions the model makes is measured by accuracy, while the model's F1 score assesses how well it balances precision and recall.

The table makes it easy to compare the performances of various models and determine which model is more effective for a particular activity. A more excellent F1 score and accuracy indicate better performance, but it is essential to consider other elements like each model's complexity and processing needs. Overall, the comparison table offers insightful information about the advantages and disadvantages of different machine learning models and can help select the most appropriate one for a given application.

Figures 3 and 4 Show the detailed comparisons of the proposed model RoBERTa with the other state-of-the-art models with box plots of their performance measures, accuracy and F1 score.

Figure 3: Comparison of evaluation metrics for Restaurant dataset

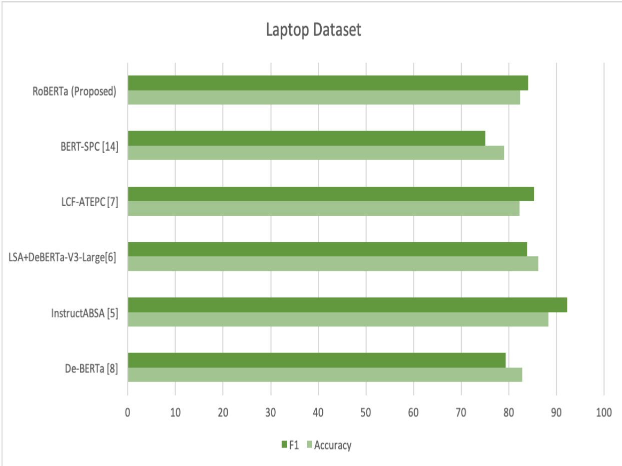

Figure 4: Comparison of evaluation metrics for laptop dataset

A classification report is a standard machine learning tool to assess how well a model performs on a classification job. A summary of different metrics, including precision, recall, F1-score, and support, is given in the report. These metrics can be used to determine how well the model is doing for each class in the classification task. A classification report plot for RoBERTa would display the model's performance for each class in a specific classification test. This would typically show each class's precision, recall, F1-score, and support values in a tabular style. These numbers show how well the model accurately identifies instances of each class and avoids false positives and negatives. Precision is the percentage of accurate positive predictions inside all optimistic forecasts for a particular class. Recall is the percentage of accurate optimistic predictions from all real positive cases for a specific class. The F1-score, a weighted precision and recall average, indicates the model's effectiveness in a specific class. The authors learned which classes the model excels at and which classes it suffers from by examining the classification report graphic.

6 Conclusion

An unimodal ABSA model uses data from a single modality, such as textual data, to assess attitudes toward specific traits or entities. Processing can be completed quickly and efficiently with this method, as it does not need to juggle many modalities, such as text and pictures. To effectively evaluate sentiment on textual data, unimodal models must incorporate language patterns, contextual cues, and dependencies within the text. As our suggested model demonstrates, sentiment prediction depends on how well the model ascertains the text's elements. Our cutting-edge model produced accuracy values of 92.35% and 82.33% for the SemEval 14 Task 4 Subtask 2 restaurant and laptop datasets, respectively. To improve results in the future, writers can apply federated learning over many state-of-the-art models or ensemble techniques to our model. The authors also intend to try the same technique on the multi-modal (Image and Text) (Image, Text and Audio) ABSA problem using an existing dataset, such as Twitter, and a real-time dataset (15 & 17).

References

[1] Gérard Biau. "Analysis of a Random Forests Model". In: *Journal of Machine Learning Research* 13.1 (2012), pp. 1063–1095.

[2] Amit Chauhan and Rajni Mohana. "Combining transfer and ensemble learning models for image and text aspect-based sentiment analysis". In: *International Journal of System Assurance Engineering and Management* (2025), pp. 1–19.

[3] Amit Chauhan, Aman Sharma, and Rajni Mohana. "A Pre-Trained Model for Aspect-based Sentiment Analysis Task: using Online Social Networking". In: *Procedia Computer Science* 233 (2024), pp. 35–44. DOI: 10.1016/j.procs.2023.12.005.

[4] Junqi Dai et al. "Does syntax matter? a strong baseline for aspect-based sentiment analysis with roberta". In: *arXiv preprint arXiv:2104.04986* (2021). DOI: 10.48550/arXiv.2104.04986.

[5] Hai Ha Do et al. "Deep learning for aspect-based sentiment analysis: a comparative review". In: *Expert systems with applications* 118 (2019), pp. 272–299. DOI: 10.1016/j.eswa.2018.10.003.

[6] Pengcheng He et al. "Deberta: Decoding-enhanced bert with disentangled attention". In: *arXiv preprint arXiv:2006.03654* (2020). DOI: 10.48550/arXiv.2006.03654.

[7] Vikramaditya Jakkula. "Tutorial on Support Vector Machine (SVM)". In: *School of EECS, Washington State University* 37.2.5 (2006), p. 3.

[8] Yingshi Jiang and Zuodong Sun. "Intelligent Music Content Generation Model Based on Multimodal Situational Sentiment Perception". In: *Informatica* 49.5 (2025). DOI: https://doi.org/10.31449/inf.v49i5.6846.

[9] Prachi Juyal. "Classification Accuracy in Sentiment Analysis using Hybrid and Ensemble Methods". In: *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*. IEEE. 2022, pp. 583–587. DOI: 10.1109/AIC55036.2022.9840020.

[10] Yue Kang et al. "Natural language processing (NLP) in management research: A literature review". In: *Journal of Management Analytics* 7.2 (2020), pp. 139–172. DOI: 10.1080/23270012.2020.1737994.

[11] Akbar Karimi, Leonardo Rossi, and Andrea Prati. "Improving bert performance for aspect-based sentiment analysis". In: *arXiv preprint arXiv:2010.11731* (2020). DOI: 10.48550/arXiv.2010.11731.

[12] D. K. Kirange, Ratnadeep R. Deshmukh, and Deepali Kirange. "Aspect based sentiment analysis SemEval-2014 Task 4". In: *Asian Journal of Computer Science and Information Technology* 4.8 (2014), pp. 72–75. DOI: 10.15520/ajcsit.v4i8.9.

[13] Yuting Luo. "Research on User Behaviour of Network Public Opinion Using Sentiment Analysis Algorithm". In: *Informatica* 48.21 (2024). DOI: https://doi.org/10.31449/inf.v48i21.6620.

[14] Shaha T Al-Otaibi and Amal A Al-Rasheed. "A review and comparative analysis of sentiment analysis techniques". In: *Informatica* 46.6 (2022), pp. 33–44. DOI: 10.31449/inf.v46i6.3991.

[15] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. "GloVe: Global Vectors for Word Representation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics. 2014, pp. 1532–1543. DOI: 10.3115/v1/D14-1162.

[16] Alexander Rietzler et al. "Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification". In: *arXiv preprint arXiv:1908.11860* (2019). DOI: 10.48550/arXiv.1908.11860.

[17] Aliea Sabir, Huda Adil Ali, and Maalim A Aljabery. "ChatGPT tweets sentiment analysis using machine learning and data classification". In: *Informatica* 48.7 (2024). DOI: https://doi.org/10.31449/inf.v48i7.5535.

[18] Kevin Scaria et al. "InstructABSA: Instruction Learning for Aspect Based Sentiment Analysis". In: *arXiv preprint arXiv:2302.08624* (2023). DOI: 10.48550/arXiv.2302.08624.

[19] Emanuel Huber da Silva and Ricardo Marcondes Marcacini. "Aspect-based sentiment analysis using BERT with disentangled attention". In: *Proceedings*. 2021. DOI: 10.48550/arXiv.2106.01237.

[20] Youwei Song et al. "Attentional Encoder Network for Targeted Sentiment Classification". In: *arXiv preprint arXiv:1902.09314* (2019). DOI: 10.48550/arXiv.1902.09314.

[21] Youwei Song et al. "Attentional encoder network for targeted sentiment classification". In: *arXiv preprint arXiv:1902.09314* (2019). DOI: 10.48550/arXiv.1902.09314.

[22] Bowen Xing and Ivor W Tsang. "Understand me, if you refer to aspect knowledge: Knowledge-aware gated recurrent memory network". In: *IEEE Transactions on Emerging Topics in Computational Intelligence* 6.5 (2022), pp. 1092–1102. DOI: 10.1109/TETCI.2022.3155731.

[23] Heng Yang and Ke Li. "Improving Implicit Sentiment Learning via Local Sentiment Aggregation". In: *arXiv e-prints* (2021), arXiv–2110. DOI: 10.48550/arXiv.2110.00000.

[24] Heng Yang et al. "A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction". In: *Neurocomputing* 419 (2021), pp. 344–356. DOI: 10.1016/j.neucom.2020.08.089.

[25] Zhilin Yang et al. "Xlnet: Generalized autoregressive pretraining for language understanding". In: *Advances in neural information processing systems* 32 (2019). DOI: 10.48550/arXiv.1906.08237.

[26] Yong Yu et al. "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures". In: *Neural Computation* 31.7 (2019), pp. 1235–1270. DOI: 10.1162/neco_a_01199.

[27] Yiming Zhang et al. "Towards Unifying the Label Space for Aspect-and Sentence-based Sentiment Analysis". In: *arXiv preprint arXiv:2203.07090* (2022). DOI: 10.48550/arXiv.2203.07090.