



# A novel framework for aspect based sentiment analysis using a hybrid BERT (HybBERT) model

Anushree Goud<sup>1</sup> · Bindu Garg<sup>1</sup>

Received: 1 May 2023 / Revised: 19 October 2023 / Accepted: 1 November 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Sentiment analysis has turned out to be a pivotal technique for fetching insights from data in textual form, and the prominent method that has emerged is aspect-based sentiment analysis, i.e., the ABSA. ABSA follows a dissection of textual content in order to associate emotions with its distinct elements. This paper reveals the efficacy of the ABSA model while exploring the different methodologies for tackling the intricate scenarios of ABSA, majorly escalating its importance. Lying amid the spectrum of techniques, transformer-based models like BERT, RoBERTa, and DistilBERT have gained substantial traction in sentiment analysis, text extraction, and natural language processing (NLP). Numerous research endeavours have covered the most important of these transformer models to enhance ABSA performance. To successfully bridge this gap between theory and practice, we brought into consideration a hybrid BERT model, which was termed HybBERT. This model blends the strengths of BERT, RoBERTa, and DistilBERT. Using data from the comprehensive Hugging Face dataset, our study meticulously processes the shared information to identify traits related to ABSA. It represents an extensive evaluation of multiple models within the ABSA framework. Each model's performance has been scrutinised and benchmarked against other models. The assessment encompasses a spectrum of evaluation metrics, which include accuracy, precision, recall, and F1-score, that provide a holistic view of performance. Our research aims to provide an important revelation: it reflects the remarkable advancement in ABSA performance, and the outcome reveals the importance of a hybrid transformer model that takes the approach beyond the depths of sentiment analysis.

**Keywords** Aspect Based Sentiment Analysis · HybBERT · RoBERTa · BERT · DistilBERT Model · Performance Evaluation

---

✉ Bindu Garg  
brgarg@bvucoep.edu.in

Anushree Goud  
anushree.anushree.goud@gmail.com

<sup>1</sup> Computer Science and Engineering Department, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India

# 1 Introduction

The creation of textual data has recently grown due to the expansion of social media platforms. The rising usage of online technology has changed how people communicate through user-generated content on e-commerce websites, social networks, blogs, etc. Following the tremendous popularity of these technologies, there has been a great interest among the researchers to explore data mining technologies for analyzing the subjective information. One such prominent research area is sentiment analysis (SA), which is used to understand user opinion and identify their view about a particular domain, website, or product. From both a commercial and academic point of view, SA is considered as an important task. Sentiment analysis is a technique that, in general, uses the polarity of the text to determine and classify the emotions and sentiment of a particular opinion or piece of user feedback. Sentiment analysis (SA) is carried out at three levels: document level, sentence level, and aspect level, to ascertain if the text from a document, a phrase, or an aspect expressed is positive, negative, or neutral. Regardless of the entities and other factors present, the majority of techniques now in use attempt to discern the polarity of the text or phrase. Contrarily, aspect-based sentiment analysis (ABSA), which is the focus of SA, aims to identify the aspects of various entities and examine the sentiment expressed for each aspect.

ABSA focuses on identifying the sentiments rather than the structure of the language. An aspect is basically related to an entity and the fundamental principle of an aspect is not restricted to any decision-making process, and can be extended towards the understanding of thoughts, user perspective, influence of social factors and ways of thinking. Hence, ABSA can be considered as an effective tool for understanding the sentiments of the users over a period of time across different domains. Because of its ability to achieve accurate sentiment classification considering different aspects, ABSA has gained huge significance in recent times.

The process involved in ABSA is categorized into three important processing stages such as; Sentiment evolution, Aspect Extraction (AE), and Aspect Sentiment Analysis (ASA) (SE). Several kinds of aspects, including explicit aspects, implicit aspects, aspect words, entities, and opinion target expressions, are retrieved in the first step. The next stage is related to the classification of sentiment polarities for a specific aspect or entity. This stage also involves the formation of interactions and analyzes the inter-relationship, contextual, and semantic relationships between multiple entities to increase sentiment categorization accuracy. The study of the user's emotions towards various components over time is related to the third stage. This comprises the user's social characteristics and experience, which are two key factors in sentiment analysis.

Compared to fundamental sentiment analysis, ABSA is a complex task since it incorporates both sentiments and aspects. The process of ABSA suffers from multifaceted challenges in terms of extracting OTE, aspects with neutral sentiments, and explicit aspects. In addition, it is also complicated to examine the relationship between various data objects for enhanced ABSA. The advancements in the field of natural language processing (NLP) with the aid of AI-based machine learning and deep learning have had a significant impact on the architecture of pre-trained models like ELMo, BERT, Roberta, and DistilBERT. These models were pre-trained using a lot of unlabelled text, and because of their versatility, ABSA performance has increased without the need for labelled data. The use of a hybrid BERT (HybBERT) model, which incorporates BERT, Roberta, and DistilBERT models for ABSA, is the main topic of this research.

The prominent contributions of this research can be summarized as follows:

- This paper provides a comprehensive evaluation of the ABSA using the proposed HybBERT model.
- The bi-directional transformer models used in this paper are pre-trained over large-scale unlabeled textual data to represent the language, which can be fine-tuned to perform specific learning-based tasks such as ABSA.
- The paper presents a case study wherein two models are combined and the performance of the hybrid model is tested against a single model.

The paper is further structured as follows: Section 2 discusses the study of various existing research methodologies on ABSA. Section 3 discusses the proposed methodology for ABSA using the HybBERT model. Section 4 discusses the results of the experimental analysis with respect to different combinations of the ABSA models, and Section 5 concludes the paper by outlining the experimental observations.

## 2 Related Works

The concept of ABSA is not a straightforward approach and incorporates a lot of challenges. Several researchers have analyzed these challenges and have come up with different solutions that have proven their efficacies in identifying and classifying different sentiments (Mercha & Benbrahim, 2023) [1] (Chandra & Jana, 2020) [2] (Gadri et al., 2022) [3] (Liu et al., 2020) [4]. These techniques have provided pictorial representations and structured approaches to handle complex tasks such as emotion recognition and sentiment analysis. Since an aspect is represented using different words, it might require more than one classification algorithm and in such cases both ML and DL models have shown promising results (Do et al., 2019) [5]. Sentiment Analysis (SA) makes use of both syntactic and semantic information for classifying the polarities (Rezaeinia et al., 2019) [6]. Here, the context of each word in the sequence might be different and the context is realized based on the other words in the sequence. Different DL models such as Convolutional Neural Network (CNN) (Liao et al., 2017) [7], Recurrent Neural Network (RNN) (Usama et al., 2020) (M. Usama, B. Ahmad, E. Song, M. S. Hossain, M. Alrashoud, and G. Muhammad) [8, 45], Long Short Term Memory (LSTM) (Gandhi et al., 2021) [9] etc. attempt to understand the context/aspect of the word with long term dependency. By studying the long-term dependencies, attempts have been made to learn and recognize the features of the word or sentence. Based on this, it can be said that it is crucial for the model to learn the aspect of words by analyzing the aspect of complete sentences without depending on the length and bidirectionality of the aspects of adjacent words in parallel (Wei Song, Zijian Wen, Zhiyong Xiao, and Soon Cheol Park) [10]. A ML based sentiment analysis for predicting different sentiments from the social media data is proposed in (Wu et al., 2020) [11] (Broek-Altenburg & Adam J.Atherly) [32] (Li, Z., Fan, Y., Jiang, B) [41]. By examining the semantic associations between two words, a hybrid strategy integrating bi-directional long short-term memory (Bi-LSTM) and convolutional neural networks (CNN) is intended to detect various labels of emotion derived from psychiatric social texts (P. Bhuvaneshwari, A. Nagaraja Rao, Y. Harold Robinson & M. N. Thippeswamy) [38] (Hengyun Li, Bruce X.B. Yu, Gang Li, Huicai Gao) [43]. Results show that the hybrid Bi-LSTM—CNN model outperforms other models compared to other sentiment analysis models. However, the performance of the model can be improved by capturing more sentiment information. A supervised ML based sentiment

analysis approach designed using Adaboost and multilayer perceptron (MLP) models is designed in (Aishwarya et al., 2019) [12]. By identifying the polarity of the available texts from the social media data, the sentiments were classified. Furthermore, the Adaboost-MLP model classified the data into multiple small classes and a novel approach of using upper-case letters and repetitive letters is used to balance the sentiment weight factors. However, a larger dataset is used to train the model and the performance is tested using small samples. For sentiment analysis and emotion recognition, a DL-based multi-task learning architecture is used (Akhtar et al., 2019) [13]. A bi-directional Gated Recurrent Unit (biGRU) model is employed for extracting the contextual data. A context level inter-modal attention technique is implemented for accurately predicting the sentiments using CMU-MOSEI dataset. Although the model achieves better performance in a multi-task learning environment, there is a need to explore other aspects such as emotion classification and intercity prediction in the proposed multitask framework. In addition, it is also important to analyze the sentiment considering the entity or aspect of the sentence. With its ability to overcome the drawback of long term dependency, LSTM models have gained huge significance and the same has been successfully implemented for ABSA (Bao et al., 2019) [14] (Xing et al., 2019) [15] (Xu et al., 2020) (B. Xing, L. Liao, D. Song, J. Wang, F. Zhang, Z. Wang, and H. Huang, 2019) [16, 46]. The effectiveness of LSTM models has been experimentally validated in several works (Al-Smadi et al., 2019) [17] employed the LSTM model for ABSA of Arabic reviews. ABSA performance increased by 39%. (Alexandridis et al., 2021) [18] implemented a knowledge based DL architecture for ABSA by integrating a hybrid Bi-LSTM with convolutional layers and an attention mechanism for enhancing the textual features. The LSTM is combined with fuzzy logic for ABSA in (Sivakumar & Uyyala, 2021) [19] for reviewing mobile phones. A highest accuracy of 96.93% was obtained when compared for other techniques. The review of DL approaches for ABSA presented in (Do et al., 2019) [5] states that the research related to ABSA and deep learning is still in the infant stage and there is a great scope to investigate further. Considering the relationship between an opinion and the aspect, the performance of the ABSA models can be improved by simultaneously extracting and classifying the sentiment, and aspect. Most of the works have focused only on aspect categorization and the models that are used to perform aspect recognition and sentiment analysis have not achieved desired performance. Hence, it is suggested to design a combined approach which can perform both tasks and provide a more precise sentiment analysis at aspect level. In this context, the implementation of pre-trained language models can be investigated for ABSA (Zhang et al., 2022) [20] (Shim et al., 2021) [21] (Troya et al., 2021) [22].

Fundamentally, the learning process of pre-trained language models is different from conventional pre-trained embedding models in terms of word embedding from a given aspect. The pre-trained language models represent the features from unsupervised text and are fine-tuned to overcome the problem of labeled data deficiency for training. Recently, transformer based methods such as BERT (Devlin et al., 2018) [23], Roberta (Wang et al., 2019) [24], DistilBERT (Sanh et al., 2019) [25], XLNet (Yang et al., 2019) [26] etc. are the most prominent bidirectionally pre-trained language models (Mao et al., 2020) [27]. The BERT model consists of several attention layers and is trained to perform downstream tasks such as emotion recognition and sentiment analysis. The BERT model can be trained to obtain deep bidirectional representations from unlabeled text by concatenating all the layers in the model (Sun et al., 2019) [28]. Hence, a pre-trained BERT model can be fine-tuned by adding one extra additional layer for performing a wide range of tasks without requiring any major architectural modifications (9Li et al., 2019) [29]. The performance of the BERT model with other pre-trained transformer models such as RoBERTa, DistilBERT, and XLNet is presented in (Adoma et al., 2020) [30] for identifying emotions from the texts. This research analyzes the sentiment by considering

the output of each model and is compared with other models. To classify various emotions, including anger, contempt, sorrow, fear, joy, humiliation, and guilt, all three models are refined using the ISEAR data. Experimental analysis shows that the RoBERTa, XLNet, BERT, and DistilBERT models achieve an accuracy of 74.31%, 72.99%, 70.09%, and 66.93% respectively. These models exhibit better accuracy as an individual model. (Phan et al., 2020) [31] extracted the syntactical features for ABSA using BERT, and RoBERTa. Results show that the proposed approach achieves phenomenal results. Although these models have shown excellent results as an individual model, there is a need for more exploration in this aspect. In future, an ensemble model can be implemented which combines all three models or any two models for improving the accuracy of emotion recognition and sentiment analysis. Considering the excellent attributes of the transformer based models, this research emphasizes the application of Bert, Roberta, and DistilBERT models for ABSA.

The aspect based sentiment analysis is a mixed approach in sentiment analysis which encapsulates multiple challenges. Researchers have successfully handles these challenges and laid out effective solutions as represented by their contributions in the classification of these sentiments (Mercha & Benbrahim, 2023) [1], (Chandra & Jana, 2020) [2], (Gadri et al., 2022) [3], and (Liu et al., 2020) [4]. These solutions utilize the visual aids and structured methodologies for managing all the intricate tasks such as emotion recognition and sentiment analysis.

Research	Contributions
Mercha & Benbrahim	Identified challenges and proposed solutions [1]
Chandra & Jana	Addressed challenges and provided insights [2]
Gadri et al	Proposed effective strategies for sentiment analysis [3]
Liu et al	Introduced innovative approaches for sentiment analysis [4]

LSTM models are know for representing long-term dependency issues which are successfully applied in ABSA (Bao et al., 2019) [14], (Xing et al., 2019) [15], and (Xu et al., 2020) [16]. These model representations have clearly demonstrated substantial performance improvements, for example, 39% enhancement is ABSA performance for Arabic reviews (Al-Smadi et al., 2019) [17].

Research	LSTM Application
Bao et al	Successful LSTM-based ABSA implementation [14]
Xing et al	Improved ABSA using LSTM models [15]
Xu et al	Application of LSTM for effective ABSA [16]

The pre trained models like BERTm Roberta, and DistilBERT have gained prominence in ABSA research (Zhang et al., 2022) [20], (Shim et al., 2021) [21], and (Troya et al., 2021) [22]. These two directional models offer the most unique advantages in the process of fine tuning with an aim to achieve exceptional results in different sentiment related tasks.

Transformer Model	Key Features & Advantages
BERT	Bidirectional contextual representations [23]
RoBERTa	Optimized pre-training for enhanced performance [24]
DistilBERT	Compact version of BERT with efficient training [25]

ABSA have multiple challenges that researchers have tried to address through different innovative solutions. DL models, LSTM approaches, and transformer models are held at great positions to advance sentiment analysis, especially within the complex contexts like ABSA.

The concept of ABSA is not a straightforward approach and incorporates a lot of challenges. Several researchers have analyzed these challenges and have come up with different solutions that have proven their efficacies in identifying and classifying different sentiments. Here, the context of each word in the sequence might be different and the context is realized based on the other words in the sequence. By studying the long-term dependencies, attempts have been made to learn and recognize the features of the word or sentence. Based on this, it can be said that it is crucial for the model to learn the aspect of words by analyzing the aspect of complete sentences without depending on the length and bidirectionality of the aspects of adjacent words in parallel. A. Nagaraja Rao, Y. Harold Robinson & M. N. Thippeswamy) [38] (Hengyun Li, Bruce X.B. Yu, Gang Li, Huicai Gao) [43]. Results show that the hybrid Bi-LSTM—CNN model outperforms other models compared to other sentiment analysis models. However, the performance of the model can be improved by capturing more sentiment information.

A bi-directional Gated Recurrent Unit (biGRU) model is employed for extracting the contextual data. A context-level inter-modal attention technique is implemented for accurately predicting sentiments using the CMU-MOSEI dataset. Although the model achieves better performance in a multi-task learning environment, there is a need to explore other aspects, such as emotion classification and intercity prediction, in the proposed multitask framework. In addition, it is also important to analyze the sentiment by considering the entity or aspect of the sentence.

Fundamentally, the learning process of pre-trained language models is different from conventional pre-trained embedding models in terms of word embedding in a given aspect. The pre-trained language models represent the features from unsupervised text and are fine-tuned to overcome the problem of labelled data deficiency for training.

The BERT model consists of several attention layers and is trained to perform downstream tasks such as emotion recognition and sentiment analysis. The BERT model can be trained to obtain deep bidirectional representations from unlabeled text by concatenating all the layers in the model (Sun et al., 2019) [28].

Although these models have shown excellent results as individual models, there is a need for more exploration in this aspect. In the future, an ensemble model can be implemented that combines all three models or any two models to improve the accuracy of emotion recognition and sentiment analysis. Considering the excellent attributes of the transformer-based models, this research emphasizes the application of the Bert, Roberta, and DistilBERT models for ABSA.

### 3 Proposed Methodology: Aspect-Based Sentiment Analysis (ABSA) using HybBERT model

In order to understand the user's emotion, this study analyses the sentiment from text data while taking into account several factors. For sentiment analysis, this study uses a hybrid BERT model known as the hybBERT model. The proposed hybrid model combines 3 bi-directional transformer based models such as Bert, Roberta, and DistilBERT for simultaneously performing aspect recognition and sentiment analysis. These models are pre-trained over large-scale unlabeled textual data to represent the language and are fine-tuned to perform ABSA. The ABSA is analyzed through different combinations of models wherein

two models are combined and the performance is compared with a single model. The work flow of the proposed approach is illustrated in Fig. 1.

## 4 Materials and Methods

From an academic and professional perspective, sentiment analysis is becoming more and more recognised as a crucial task. However, the bulk of current systems aim to identify the overall polarity of a sentence, paragraph, or text span, regardless of the entities (such as laptops and restaurants) and their components (such as the battery, screen, food, and service). Aspect-based sentiment analysis (ABSA), in contrast, is the focus of this task. Its objective is to detect the aspects of the target entities and the sentiment expressed towards each aspect. Datasets of customer reviews with human-authored annotations indicating the specified properties of the target entities and the polarity of each property's sentiment will be made available.

The task specifically consists of the following subtasks:

### Subtask 1: Extracting aspect terms

Determine the aspect terms that are present in a set of sentences with pre-identified entities (such as restaurants) and provide a list of all the different aspect terms. A specific aspect of the target entity is identified by an aspect word.

For instance, "The food was nothing special, but I loved the staff," or "I liked the service and the staff, but not the food. The phrase "The hard disc is very noisy" is the only instance of an aspect term that consists of more than one word (for example, "hard disc").

### Subtask 2: Polarity of the aspect word

Determine the polarity of each aspect term in a statement, such as whether it is positive, negative, neutral, or in conflict (both positive and negative).

Here is the dataset link which will be used in the code:

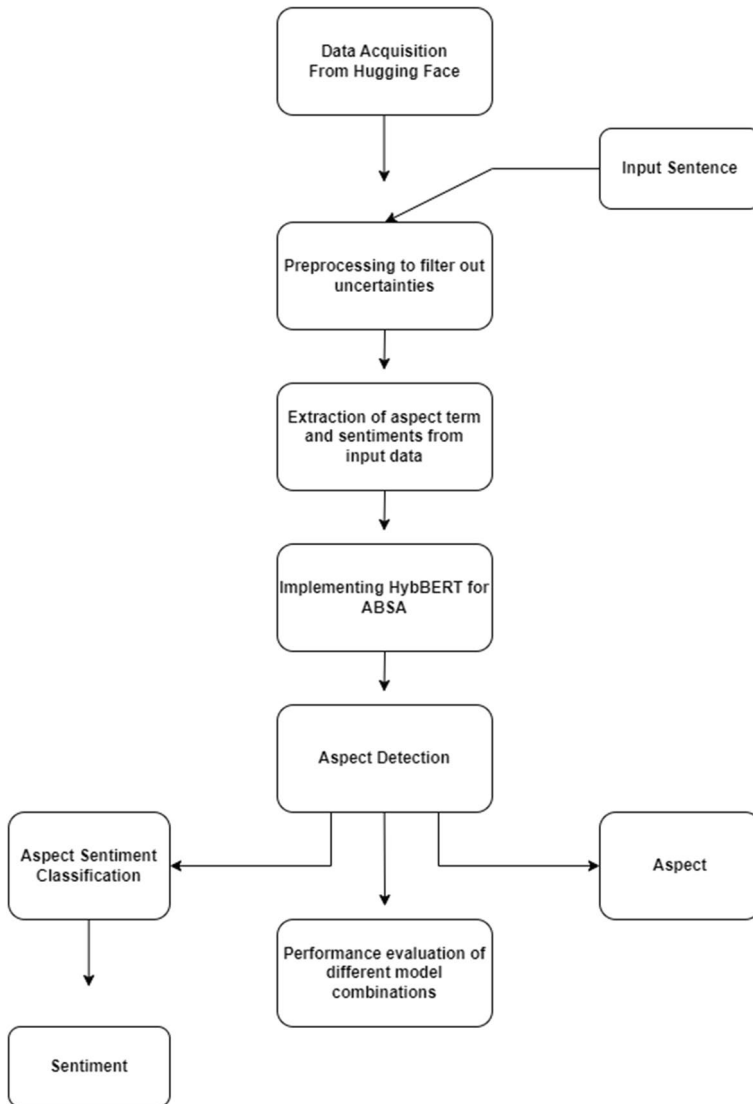
<https://huggingface.co/datasets/Yaxin/SemEval2014Task4Raw/viewer/All/train>

## 4.1 Data Acquisition

The first part of the process involves gathering input data from the database, which is known as data collection or acquisition. In this research, the data is collected from the Hugging Face dataset (Geetha & Renuka, 2021) [42]. The dataset features define the internal structure of the dataset and contain high-level information regarding different fields such as the number of columns, types, data labels, and conversion methods. The dataset consists of data labels such as text, aspect terms, aspect categories, domain and sentence Id which helps to predefine the set of classes and are stored in the form of integers.

## 4.2 Data Preprocessing

The data extracted from the dataset is in basic raw format and needs to be processed before feeding it to the model for training. By removing uncertainties like missing data, null values, noisy data, and other irregularities, the raw data is transformed into a clean dataset at this step. Preprocessing is the basic step in the ABSA process and involves different steps such as:



**Fig. 1** Workflow of the proposed ABSA

- Sorting the dataset according to the column
- Shuffling the dataset
- Filtering the rows according to the index
- Concatenating the datasets with the same column types

The obtained data consisted of multiple columns containing responses, opinions, and feedback. The columns with sentiment labels and individual response are considered in this research for ABSA. Hence, these two columns are subjected to preprocessing. It was observed that few columns consisted of sentiment labels without any textual responses. In



addition, the data also consisted of special characters, irrelevant tags, double spacing and redundant expressions, which are also filtered since they can have a negative impact on the performance of sentiment analysis. The dataset, with a total of 4631 rows, is taken into consideration for the analysis, with a total of 2 columns after filtering. The data are divided into training, testing, and validation samples, with 3509 samples being used for training, 1028 samples being used for testing, and 94 samples being used for validation.

### 4.3 Aspect-based feature extraction

In general, feature extraction is performed extracting relevant features for the sentiment analysis and classification process (Meng et al., 2019) [33]. When the majority of the characteristics are not enough to contribute to the total variance, the dimensionality is reduced by extracting just the relevant features. The computing time will be cut down, and the overall performance will be improved, by removing unnecessary and duplicate features. Aspect-based feature extraction techniques extract the essential features that are crucial for defining the text's primary traits, or aspects. This helps in finding important textual elements that reflect an opinion or a mood. In order to choose pertinent and appropriate features from the text, new feature subsets are acquired at this stage. For aspect-based feature extraction, two bags of words are created, where the first group contains aspects and the second group contains the tendency of sentiment polarity.

In aspect based feature extraction, it is highly challenging to extract explicit aspects, aspect categories, and Opinion Term Extraction (OTE) based on which the sentiment is expressed (Zhang et al., 2022) [34]. An explicit aspect here means that the sentiment is present in the text, for instance: "the food in this restaurant is great" is an example of an explicit aspect where food is an explicit aspect and restaurant is the explicit entity. While performing aspect extraction, the relation between different aspects is also analyzed for identifying the coherent, consistent, and aspects with high similarity from the dataset in order to improve the overall representation of the extracted aspects.

### 4.4 Aspect Classification using HybBERT model

This study implements a HybBERT model for aspect classification and ABSA. During the training phase, the models will use examples from the data set to determine how to transform the input text into a class of pertinent attributes (Lu Xu, Lidong Bing, Wei Lu, Fei Huang, November 2020) [25]. The feature vectors and appropriate class labels will be given for the proposed HybBERT model. The classification models used for ABSA are explained as follows:

#### 4.4.1 BERT Model

BERT is a bi-directional transformer model which is used to pre-train a large volume of unlabeled textual data to learn how to represent a language and to fine tune the model for performing a specific task. BERT model outperforms when applied for natural language processing (NLP) and sentiment analysis tasks (Howard & Ruder, 2018) [35] (Radford et al., 2018) [36]. The improved performance of the BERT model is due to the bidirectional transformer and its ability to pre-train the model for predicting the next sentence. BERT employs a fine tuning mechanism which almost eliminates the need for a specific architecture for performing each task. Hence, the BERT model is considered as an intelligence model which can reduce the necessity of prior knowledge before designing the model and

instead it enables the model to learn from the available data. The BERT model has two architectural settings which are the BERTBASE and BERTLARGE. The BERTBASE has 12 layers with 768 hidden dimensions and 12 attention heads (in transformer) with 110 M number of parameters. The BERTLARGE, on the other hand, has 24 layers, 1024 hidden dimensions, and 16 attention heads (in a transformer) with 340 M parameters.

In this research the BERT model is fine-tuned to perform two tasks namely aspect extraction (AE) and ABSA. For aspect extraction, the BERT model is provided with continuous samples of data which are labeled as A and B as aspects. The input text 'n' words are constructed as  $x=(x_1, \dots, x_m)$ . After  $h=BERT(x)$ , a dense layer along with a Softmax layer is applied for each position of the text denoted as  $l=\text{softmax}(W * h+b)$  where, W denotes the text in the word, l is the length of the sequence, h and b are the dimensions. Softmax is applied along with the dimension of labels for each text and the labels are predicted based on the position of the text.

For aspect classification, the BERT model is fine tuned to classify the polarity of the sentiment (Positive, negative, or neutral) based on the aspect extracted from the text. For ABSA, the model is provided with two inputs namely an aspect and a text defining the aspect. Let  $x=(p_1, \dots, p_m)$  be an aspect with 'n' number of tokens, and  $x=(t_1, \dots, t_m)$  defining the text containing the aspect.

Similar to aspect extraction, after  $h=BERT(x)$ , the learning representation of the BERT model is leveraged to obtain the polarity of the text, based on which the polarity of the text is calculated.

The fine tuning of the BERT model is simple and straightforward since the learning capability of the transformer supports the model in various down streaming tasks whether in terms of classifying the single text or multiple texts by interchanging the respective inputs and outputs (Azhar & Khodra, 2020) [37]. For classifying the sentiments in multiple text pairs, the BERT model derives a common pattern which can automatically encode the text pairs and performs bidirectional cross attention between two sentences.

#### 4.4.2 ROBERTA Model

An improved version of the BERT model, the ROBERTA model was first presented by Facebook. The ROBERTA model is obtained as a retrained BERT model with more computing capabilities and better training process. In order to achieve enhanced training, ROBERTA eliminates the Next Sentence Prediction (NSP) task from the existing BERT model and employs a dynamic masking approach such that the masked token can be changed while training the model for several epochs. This factor also enables the ROBERTA model to train using large training samples. The text obtained from the dataset is given as input to the model, and the outputs are obtained from the last layer of the model. While using the ROBERTA model for ABSA, the model is fine-tuned using a cross-entropy loss function, which enables the model to effectively leverage its potential in terms of ABSA. Another important advantage of the ROBERTA model is utilization of dynamic masking, wherein the model is provided with a dynamic set of data samples from the text instead of providing only a fixed set of data samples as in the BERT model. This improves the learning ability of the ROBERTA model by enabling it to learn from a diversified set of data. In addition, the dynamic masking also makes the model more resilient to the changes in the input data, which is more important while handling diverse, irregular and inconsistent text. In addition, the ROBERTA model also employs a No Mask Left Behind (NMLB) technique which makes sure that all the data samples are masked at least once during training unlike BERT models which uses a Masked Language Modeling (MLM) technique that masks only 15% of the data samples. This improves the representation ability of the text and helps the model to exhibit better ABSA performance. For ABSA, the ROBERTA model consists of four modules namely, (i) a word embedding layer, (ii) a semantic representation layer, (iii) cross attention layer, and (iv) a classification output layer.

In the word embedding layer, the ROBERTA model is pre-trained using multiple embedded layers of sentences and aspects to obtain an improved representation of aspects. The transformer used in the ROBERTA model helps in capturing the bidirectional relationship between the sentences and helps in mapping each word  $w_i$  in the sentence and its aspect to a low dimensional vector  $v_i \in \mathbb{R}^{d_w}$ , where  $d_w$  is the dimension on the word vector. The pre-training of the model results in the sentences and aspects that caters the sentence vector  $\{v_1s, v_2s, v_3s, \dots, vns\} \in \mathbb{R}^{n \times d_w}$  and the corresponding aspect vector  $\{v_1s, v_2s, v_3s, \dots, vns\} \in \mathbb{R}^{m \times d_w}$ . Both the sentence and the aspect vectors obtained from the word embedding layer are embedded with the word vector and are given as input to the semantic representation layer. In the cross attention layer, the word vectors obtained from the semantic representation layer are used to determine the effect of weight of the aspect and sentence. Initially, the cross attention layer takes the aspect and sentence denoted as  $h_a \in \mathbb{R}^{n \times d_w}$  and  $h_s \in \mathbb{R}^{n \times d_w}$  respectively and later the matching matrix  $I = h_s \times h_a$  is calculated. Further, a Softmax function is applied to the matrix  $I$  to obtain a degree  $\alpha$  and  $\beta$  of the sentence to the aspect. Lastly, the average of the attention is calculated using the attention degree  $\beta$  of the sentence to the aspect is calculated. The most important aspect is represented using term  $\beta^*$  which along with degree  $\alpha$  is used to calculate the final output denoted as  $\gamma \in \mathbb{R}^m$ , as shown in below equations:

$$\alpha_{ij} = \exp(I_{ij}) / \sum_i (I_{ij}) \quad (1)$$

$$\beta_{ij} = \exp(I_{ij}) / \sum_i (I_{ij}) \quad (2)$$

Correspondingly, the important term of aspect  $\beta^*$  is represented as follows:

$$\beta^* = (1/n) \sum (\beta) \quad (3)$$

Finally, the impact is denoted as

$$\gamma = \alpha^* \beta^* \quad (4)$$

The classification output layer generates the output of the model by convoluting the results of the semantic representation layer and uses it to predict the final polarity of the sentiment corresponding to the aspect, as shown in below equations:

$$r = hw^* \gamma \quad (5)$$

$$p = \text{softmax}(w^* r + b) \quad (6)$$

The term 'p' defines the probability distribution of sentiment analysis, w and b define the weight and bias of the matrices respectively. The model is trained using a cross-entropy loss function and L1 regularization, as shown in Eq. 7.

$$L1 = -\sum_i \sum_c \in C [((y_i = c) * \log(p(y_i = c))) + \lambda ||\theta||_2] \quad (7)$$

wherein  $p(y_i = c)$  is the predicted sentiment polarity and  $(y_i = c)$  is the actual sentiment polarity, and  $\lambda$  is the regularization parameter.

#### 4.4.3 DISTILBERT Model

The DISTILBERT model is a distilled version of the BERT model that exhibits almost the same performance as the BERT model but uses only half the number of parameters. In

particular, the DISTILBERT model does not incorporate any word-based embeddings and makes use of only half of the layers from the BERT model. The main difference between the BERT and DISTILBERT models are illustrated in Table 1.

The distillation mechanism used by the DISTILBERT model approximates the large neural network model by using a smaller one. The main idea behind this model is to train a larger neural network once and the output distributions can be approximated using a smaller model. DISTILBERT models are trained using a triple loss function which is a linear combination of the distillation loss ( $L_{ce}$ ) with other loss functions such as training loss, MLM loss (LMLM), and a cosine embedding loss ( $L_{cos}$ ) which establishes a coordination between the aspect and sentence using the state vectors. The distillation loss is determined as follows:

$$L_{ce} = \sum t_i * \log(s_i) \quad (8)$$

where,  $t_i$  ( $s_i$ ) is the probability of the estimation.

Using the triple loss function, a 40% smaller transformer (Vaswani et al., 2017) [38] can be pre-trained using distillation with 60% faster inference time. Studies have shown that the triple loss function parameters have the capacity to achieve excellent performance (Sanh et al., 2019) [39]. This research uses a pre-trained DISTILBERT model from the hugging face transformer library for ABSA, which is distilled from the fundamental BERT model. The excellent performance of the DISTILBERT model is due to its ability to extract bidirectional contextual information from the text data during the training process. Besides, the compactness of the DISTILBERT model increases the scalability of the model and helps them achieve better performance in real-time sentiment analysis tasks (Kevin Scaria, Himanshu Gupta, Siddharth Goyal, Saurabh Arjun Sawant Swaroop) (Lu Xu, Lidong Bing, Wei Lu, Fei Huang) [44, 46] while retaining 97% of the BERT's performance. Figure 2 shows a schematic illustration of the three models utilized in the ABSA method.

Figure 2 is a schematic illustration of ABSA Models with BERT, RoBERTa, and DistilBERT that highlights their integral roles. These models have played an instrumental role in this study. The Aspect Extraction Model, which has been represented in blue is reflected to harness the power of BERT. BERT has excelled in understanding the context where aspects or features are mentioned within the textual data with the extraction of these aspects and providing an understanding of what we have discussed.

The Sentiment classification model has leveraged RoBERTa which makes it a robust variant of BERT. excelling significantly at sentiment classification which is detrimental in classifying it as positive, negative, or neutral which assures accuracy in sentiment analysis.

The contextual analysis model apparently utilizes the DistilBERT model which is a distilled version of BERT. This efficiently processes and analyzes the relationship between aspects and sentiments within a relatable broader context of the text ensuring the correct attributions of sentiments while considering contextual nuances and linguistic intricacies.

**Table 1** Difference between BERT and DISTILBERT models

Parameters used in training	BERT model	DISTIL-BERT model
The number of layers in the transformer blocks	12	6
Hidden units	768	768
Number of heads for self-attention	12	12
Total number of variables that can be trained	110 M	66 M

Together these models synergize while offering a comprehensive approach to ABSA, BERT, RoBERTa, and DistilBERT’s unique capabilities enabling nuanced and context-aware sentiment analysis while providing valuable insights into the sentiment associated with specific aspects as discussed in textual data.

The feature extractor is utilized in the ABSA technique to extract feature vectors from a brand-new text that has never been seen before. The model then generates the results using these feature vectors. Each text in the dataset has one of three sentiments: positive, negative, or neutral (Kevin Scaria, Himanshu Gupta Siddharth Goyal, Saurabh, Arjun Sawant Swaroop Mishra Chitta Baral [44]. The sentiment score, which is based on positive and negative emotions and is represented by the equation below, may be used to assess each text’s sentiment. (Ruz and others, 2020) [40].

$$\text{Sentiment score(SC)} = (\text{positive} - \text{negative}) / (\text{positive} + \text{negative} + 2) \tag{9}$$

As a result, the positive and negative describe the total number of positive and negative words in the text, and the SC may be determined using a discrete 2-valued variable C that also defines the emotion class and falls between -1 and 1, i.e.,  $C \in \{-1, 1\}$ .

The term ‘C’ records the sentiment values and their respective classes. In certain cases, it is difficult to identify the aspects from the sentiments based on the polarity values and in such cases, certain constraints (as shown in the Eq. 10) need to be followed to identify whether the sentiment is positive or negative.

$$\begin{aligned} C &= 1(\text{positive tweet}) \text{ if Sentiment score} \geq 0.1 \\ &= -1(\text{negative tweet}) \text{ if Sentiment score} < 0.1. \end{aligned} \tag{10}$$

The polarity of texts that have aspects is assessed when they are used as input to the models to determine if the text is positive, negative, or neutral. The results enable the identification and classification of the text’s sentiment. The pseudocode for this process is given as follows:

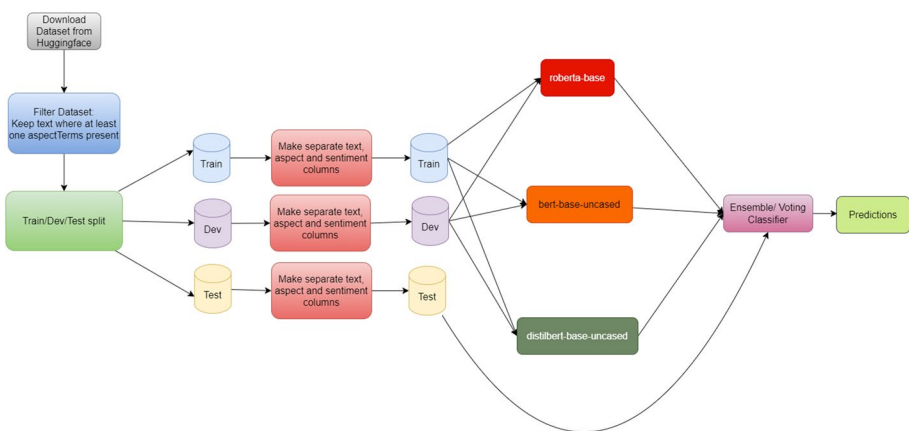


Fig. 2 Process of ABSA using different transformer model combinations

**Pseudocode**

**Input :** Labeled Text from Hugging Face Dataset

**Step 1: Preprocessing:**

Removal of redundant words  
 Identifying missing values  
 Sorting, Shuffling and Filtering the dataset

**Step 2: Aspect based Extraction of textual data:****Step 3: Extraction of explicit aspects from extracted data:****Step 4: Combine step 1 and 2:**

Pre-processed file =file path  
 name Stop word file list=path  
 name of file  
 Extracted text=file path of extracted text list

**Step 5: Aspect classification and Sentiment analysis**

For each words in adjective list  
 Extract adjectives in the text  
 data () Calculate the sentiment  
 score  
 For every positive  
 words: 1 For every  
 negative words: -1  
 Find  
 If ( $C \geq 0.1$ )  
 Classi  
 fy Positive  
 Else if (C  
 < 0.1)  
 Classif  
 y Negative  
 Else  
 Classify Neutral

**Output:** Sentiment polarity of the text data.

Most of the studies on ABSA have focused on the implementation of the transformer models as an individual model. This research emphasizes the analysis of three different combinations of the transformer model such as the BERT—ROBERTA model, BERT—DISTILBERT model, and ROBERTA—DISTILBERT model. The performance of these model combinations are discussed in the next section.

## 5 Results Analysis and Discussion

In Aspect-Based Sentiment Analysis (ABSA), the primary focus is on extracting and analysing sentiment expressed towards specific aspects or features within a piece of text, such as a customer review. To perform ABSA effectively, we need to consider several key aspects of the analysis process:

**Aspect Identification and Extraction** This is the foundation of ABSA. Identifying and extracting the aspects or features being discussed in the text is crucial. Aspects can vary depending on the domain and context. Techniques for aspect extraction can include rule-based methods, dependency parsing, or utilising pre-trained models.

**Sentiment Polarity Detection** After extracting aspects, we need to determine the sentiment polarity associated with each aspect mentioned in the text. Sentiment polarity typically falls into categories like positive, negative, neutral, or more fine-grained scales. This step often involves sentiment lexicons, machine learning models, or deep learning approaches for sentiment classification.

Based on a variety of performance criteria, including F1 Score, Recall, Accuracy, and Support, the suggested BERT—ROBERTA model, BERT—DISTILBERT model, and ROBERTA—DISTILBERT model's performance is assessed. The sections below show the simulation results for sentimental analysis, and the confusion matrix was used to gauge how effective the suggested strategy is. The performance of the classifier is assessed using a confusion matrix employing the four elements TP, FP, TN, and FN. The phrases True positive, False positive, False negative, and True negative are TP, FP, FN, and TN, respectively. An illustration of the confusion matrix is given in Fig. 3

Fig. 3 Confusion matrix

		Actual (True) Values	
		Positive	Negative
Predicted Values	Positive	TP	FP
	Negative	FN	TN

Here, TP defines the number of correctly identified positive sentiments, TN defines the number of correctly identified negative sentiments, FP is the inaccurately identified positive sentiments, and FN is the inaccurately identified negative sentiments. The expression of the different performance metrics are given in the below expressions as follows:

- *Accuracy:*

$$\begin{aligned} \text{Accuracy} &= \frac{(\text{No of correctly identified sentiments})}{(\text{Total number of sentiments in the test dataset})} \\ \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \end{aligned} \quad (11)$$

- *Precision:*

Precision is defined as the ratio of the correctly identified positive sentiments which are relevant and is given as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

- *Recall:*

Recall for a function is determined as the ratio of the correctly identified positive sentiments to the sum of true positive and false positive. The expression for recall is given as shown in equation 13.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

## 5.1 F1 Score

F1 score is also termed as an F measure score which is used to determine the harmonic mean of the precision and recall. F1 score is closely related to the accuracy of the classifier whose value lies between 0 and 1, where 0 and 1 represents the worst and best values respectively. F1 score can also be used to measure the performance of the model wherein the datasets are highly imbalanced since it uses both recall and precision to obtain an optimal value. Even if there are fewer positive samples compared to negative samples, the expression used to measure the F1 score will weigh the metric value. Correspondingly, F1 score is defined as:

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (14)$$

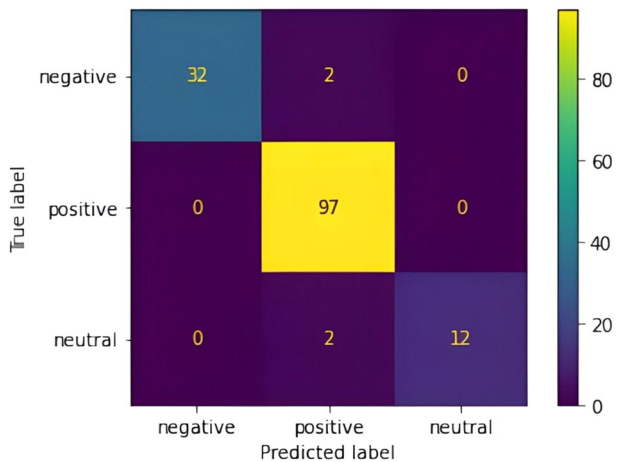
## 5.2 Performance of BERT, ROBERTA, and DISTILBERT models

Initially, the performance of the BERT, ROBERTA, and DISTILBERT models is evaluated as individually and the classification report of the respective models are discussed as follows:



**Table 2** Classification report for sentiment analysis using the BERT model

	Precision	Recall	F1 score	Support
Negative	1.00	0.94	0.97	34
Positive	0.96	1.00	0.98	97
Neutral	1.00	0.86	0.92	14
Accuracy	-	-	0.97	145
Macro Avg	0.99	0.93	0.96	145
Weighted Avg	0.97	0.97	0.97	145

**Fig. 4** Confusion matrix for the BERT model

### • Classification Performance of BERT model

The performance of the BERT model is shown in Table 2 and the corresponding confusion matrix is shown in Figure 4.

This figure is more like a report card representation for the BERT model that we have made use in the ABSA. It is a representation of how efficiently the model has performed in the classification of sentiments like the positives or negatives for different aspects of the text. This chart is a better way for us to analyze how well a BERT model is at understanding sentiment. It also reveals where the model is performing well and where an improvement is needed in ABSA something very similar to checking how many answers on a test have been right or wrong.

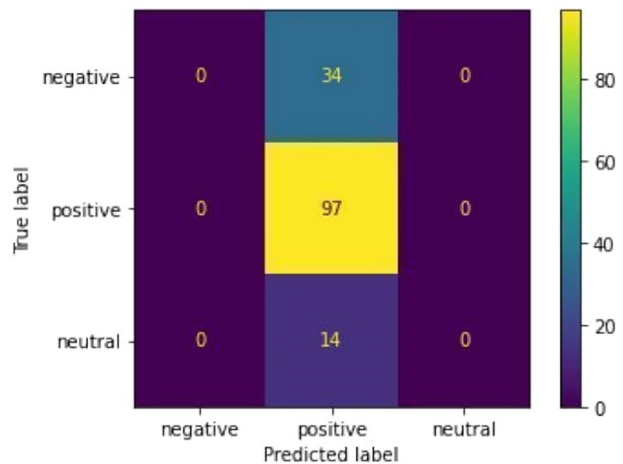
Precision also referred to as the Positive Predictive Value, indicates the measure of the proportion of appropriately defined positive instances out of all the total instances declared as positive for each class. A high precision is an indication of all the instances predicted as positive and so on for each class. A high precision is an indication of instances are classified as positive and likely to be positive.

Recall is also referred to as Sensitivity or True Positive Rate that is a measurement of the proportion that currently predicts positive instances out of all the actual instances. A high recall indicates that the model effectively captures most of the positive instances.

Support is a precise indication of the number of instances prescribed in each class that calculates the metrics. It is an understanding towards relative prevalence in the dataset.

**Table 3** Classification report for sentiment analysis using the ROBERTA model

	Precision	Recall	F1 score	Support
Negative	0.00	0.00	0.00	34
Positive	0.67	1.00	0.80	97
Neutral	0.00	0.00	0.00	14
Accuracy	-	-	0.67	145
Macro Avg	0.22	0.33	0.27	145
Weighted Avg	0.45	0.67	0.54	145

**Fig. 5** Confusion matrix for the ROBERTA model

Accuracy is the measurement of the proportion of aptly predicted instances from all the instances. It is a common metric for evaluation the entire performance of a model.

The represented information makes the BERT model to be aptly performing in the task of sentiment analysis, with high accuracy, precision, recall, and F1 score for each class. The results of the classification shows that the BERT model exhibits a phenomenal accuracy of 97.24%, F1 score of 97.19%, precision of 97.35%. And recall of 94.24%.

- **Classification Performance of ROBERTA model**

The performance of the ROBERTA model is shown in Table 3 and the corresponding confusion matrix is shown in Figure 5.

- **Classification Performance of DISTILBERT model**

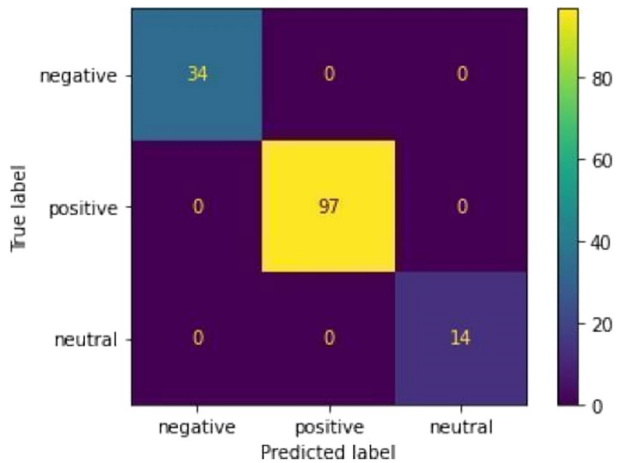
The classification report of the DISTILBERT model is shown in Table 4 and the corresponding confusion matrix is shown in Figure 6.

The DISTILBERT model achieves an outstanding classification performance by achieving 100% in terms of all evaluation metrics. In comparison to the BERT model and ROBERTA model the DISTILBERT model outperforms with an average accuracy of 100% and average precision of precision, recall, and F1 score of 100% respectively. The results show that the DISTILBERT model is more appropriate for performing real time ABSA and in

**Table 4** Classification report for sentiment analysis using the DISTILBERT model

	Precision	Recall	F1 score	Support
Negative	1.00	1.00	1.00	34
Positive	1.00	1.00	1.00	97
Neutral	1.00	1.00	1.00	14
Accuracy	-	-	1.00	145
Macro Avg	1.00	1.00	1.00	145
Weighted Avg	1.00	1.00	1.00	145

**Fig. 6** Confusion matrix for the DISTILBERT model



terms of performance, BERT can be opted as second best model and ROBERTA is the least preferred.

**5.3 Performance of different combination of transformer models**

This section discusses the performance of different combinations of transformer models such as BERT - ROBERTA model, BERT - DISTILBERT model, and ROBERTA – DISTILBERT models.

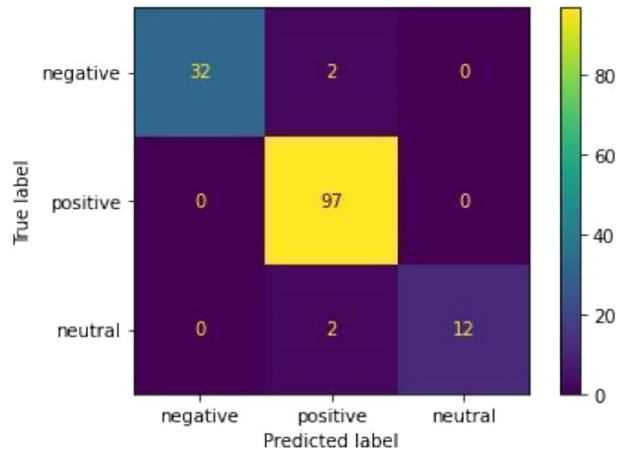
• *Performance of final model for ABSA*

The ABSA performance of the final model is shown in Table 2 and the corresponding confusion matrix is shown in Figure 7

The results of the classification shows that the final model achieves better classification performance with an accuracy of 97%, F1 score of 97%, precision of 97% and a recall of 97%.

• *Combining DISTILBERT model and ROBERTA model*

The performance of the DISTILBERT model and ROBERTA model is shown in Tables 5 and 6 and the corresponding confusion matrix for ABSA is shown in Figure 8.

**Fig. 7** Confusion matrix for the final model**Table 5** Classification report for sentiment analysis using the final model

	Precision	Recall	F1-score	Support
Negative	1.00	0.94	0.97	34
Positive	0.96	1.00	0.98	97
Neutral	1.00	0.86	0.92	14
Accuracy	-	-	0.97	145
Macro Avg	0.99	0.93	0.96	145
Weighted Avg	0.97	0.97	0.97	145

**Table 6** Classification report for sentiment analysis using the DISTILBERT—ROBERTA model

	Precision	Recall	F1 score	Support
Negative	1.00	1.00	1.00	34
Positive	1.00	1.00	1.00	97
Neutral	1.00	1.00	1.00	14
Accuracy	-	-	1.00	145
Macro Avg	1.00	1.00	1.00	145
Weighted Avg	1.00	1.00	1.00	145

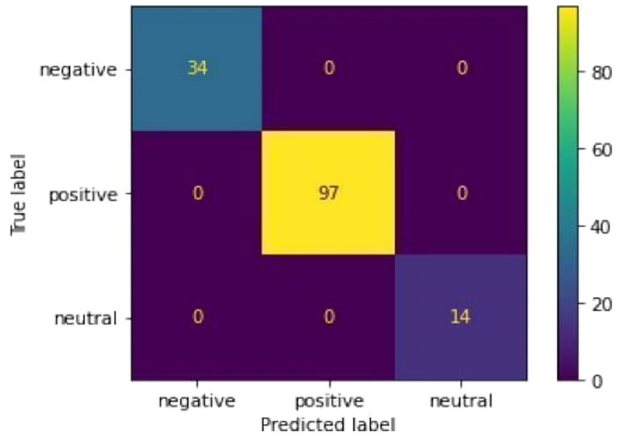
As inferred from the results, the excellent performance of the DISTILBERT - ROBERTA model validates its adaptability in real time ABSA tasks. This combination achieves 100% accuracy along with other performance evaluation metrics.

- **Combining DISTILBERT model and BERT model**

The performance of the DISTILBERT model and BERT model is shown in Table 7 and the corresponding confusion matrix for ABSA is shown in Figure 9.

Similar to the DISTILBERT - ROBERTA model, the combination of DISTILBERT - BERT model also achieves 100% accuracy along with other performance evaluation

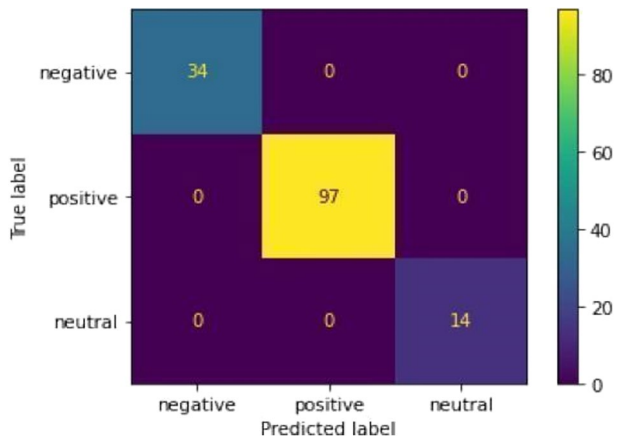
**Fig. 8** Confusion matrix for the DISTILBERT—ROBERTA model



**Table 7** Classification report for sentiment analysis using the DISTILBERT—BERT model

	Precision	Recall	F1 score	Support
Negative	1.00	1.00	1.00	34
Positive	1.00	1.00	1.00	97
Neutral	1.00	1.00	1.00	14
Accuracy	-	-	1.00	145
Macro Avg	1.00	1.00	1.00	145
Weighted Avg	1.00	1.00	1.00	145

**Fig. 9** Confusion matrix for the DISTILBERT—BERT model

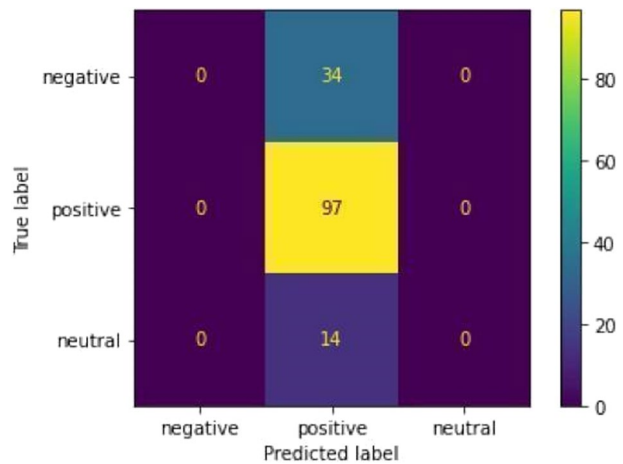


metrics. This shows that the DISTILBERT model combined with any other model can help in achieving phenomenal results.

- **Combining ROBERTA model and BERT model**

**Table 8** Classification report for sentiment analysis using the ROBERTA—BERT model

	Precision	Recall	F1 score	Support
Negative	0.00	0.00	0.00	34
Positive	0.67	1.00	0.80	97
Neutral	0.00	0.00	0.00	14
Accuracy	-	-	0.67	145
Macro Avg	0.22	0.33	0.27	145
Weighted Avg	0.45	0.67	0.54	145

**Fig. 10** Confusion matrix for the ROBERTA—BERT model**Table 9** Classification report for sentiment analysis using the combined dual model prediction

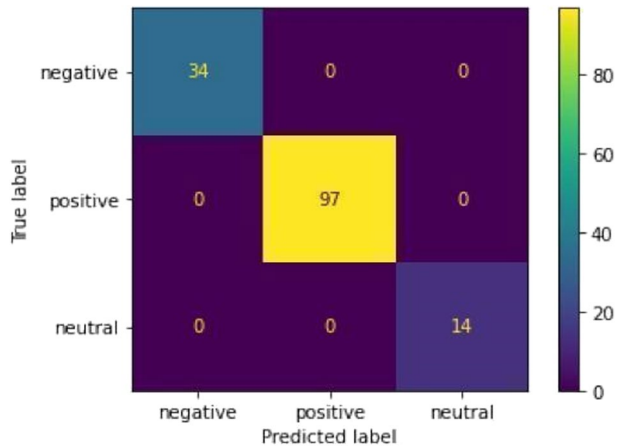
	Precision	Recall	F1 score	Support
Negative	1.00	1.00	1.00	34
Positive	1.00	1.00	1.00	97
Neutral	1.00	1.00	1.00	14
Accuracy	-	-	1.00	145
Macro Avg	1.00	1.00	1.00	145
Weighted Avg	1.00	1.00	1.00	145

The performance of the ROBERTA - BERT model is shown in Table 8 and the corresponding confusion matrix for ABSA is shown in Figure 10.

The classification report of the ROBERTA - BERT model is shown in Table 8. The combined ROBERTA - BERT model achieves a moderate accuracy of 67% along with other performance evaluation metrics. Results show that the performance of the ROBERTA model underperforms when used as an individual classifier or combined with other models.

- *Performance of the combined dual model predictions*

**Fig. 11** Confusion matrix for the combined dual model prediction



The performance of the combined dual model predictions is shown in Table 9 and the corresponding confusion matrix for ABSA is shown in Figure 11

The classification report of the combined dual model is shown in Table 9. Similar to the DISTILBERT model and DISTILBERT- BERT model, the combined dual model achieves an excellent accuracy of 100% along with other performance evaluation metrics.

The confusion matrix obtained after experimental analysis is presented from Figures 4, 5, 6, 7, 8, 9, 10 and 11 for BERT, ROBERTA, and DISTILBERT models as individual models and the combined DISTILBERT- BERT, BERT - ROBERTA, and DISTILBERT-ROBERTA models along with final model and combined dual model. The classification report is generated for different candidate models for validation data.

## 5.4 Comparative Analysis

This section discusses the comparative analysis of all the transformer models with different combinations. Here, the performance is compared with respect to different case studies as discussed below:

**Case i: BERT- ROBERTA model:** Here, the performance of the combined Bert-Roberta model is evaluated, and the results are compared with those of the Distilbert model.

**Case ii: BERT- DISTILBERT model:** Here, the performance of the combined Bert-Distilbert model is evaluated, and the results are compared with those of the ROBERTA model.

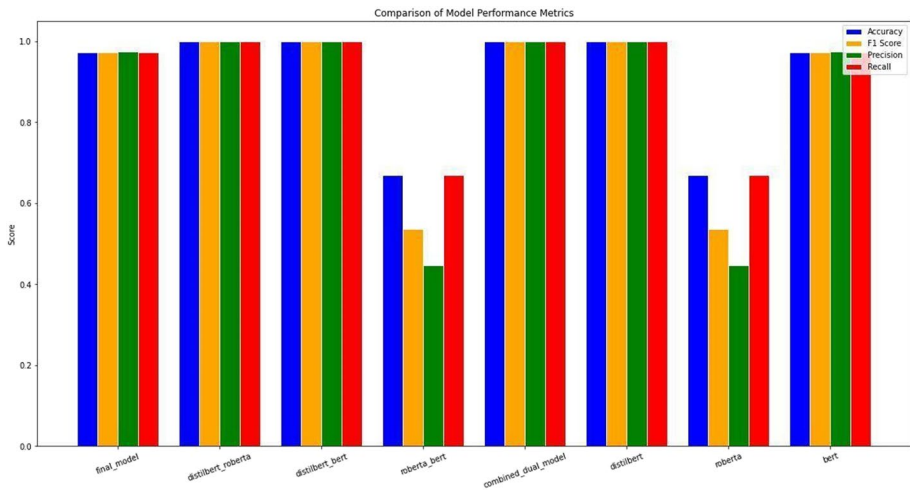
**Case iii: DISTILBERT- ROBERTA model:** Here, the performance of the combined DISTILBERT-ROBERTA model is evaluated, and the results are compared with the BERT model.

### Case 1: Combination of all models

The classification report of all combinations of the models are compared and the same is tabulated in Table 10 and the graphical representation of the ABSA performance is illustrated in Figure 12.

**Table 10** Classification report of all model combinations

Model	Accuracy	F1 score	Precision	Recall
Final Model	97.24%	97.19%	97.35%	97.24%
DISTILBERT—ROBERTA	100%	100%	100%	100%
DISTILBERT—BERT	100%	100%	100%	100%
ROBERTA—BERT	66.89%	56.32%	44.75%	66.89%
Combined Dual Model	100%	100%	100%	100%
DISTILBERT	100%	100%	100%	100%
ROBERTA	66.89%	53.62%	44.75%	66.89%
BERT	97.24%	97.19%	97.35%	97.24%

**Fig. 12** Classification performance of all model combinations**Table 11** Classification report of individual models with DISTILBERT + ROBERTA + BERT

Model	Accuracy	F1 score	Precision	Recall
Final Model	97.24%	97.19%	97.35%	97.24%
DISTILBERT	100%	100%	100%	100%
ROBERTA	66.89%	53.62%	44.75%	66.89%
BERT	97.24%	97.19%	97.35%	97.24%

### Case 2: DISTILBERT, ROBERTA, BERT, with DISTILBERT + ROBERTA + BERT

The classification performance of the individual models with combined models is shown in Table 11 and Figure 13.

### Case 3: DISTILBERT, ROBERTA with DISTILBERT + ROBERTA



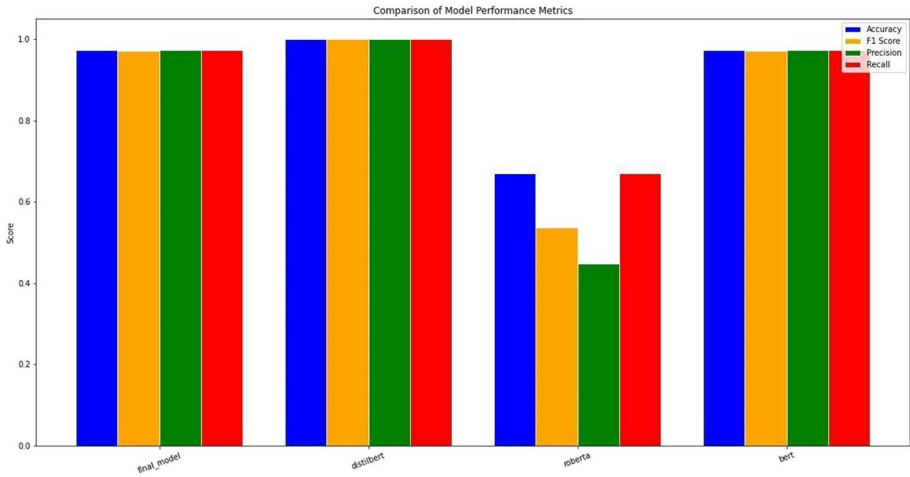


Fig. 13 Classification performance of individual models with DISTILBERT + ROBERTA + BERT

Table 12 Classification report of DISTILBERT, ROBERTA models with DISTILBERT + ROBERTA

Model	Accuracy	F1 score	Precision	Recall
DISTILBERT + ROBERTA	100%	100%	100%	100%
DISTILBERT	100%	100%	100%	100%
ROBERTA	66.89%	53.62%	44.75%	66.89%

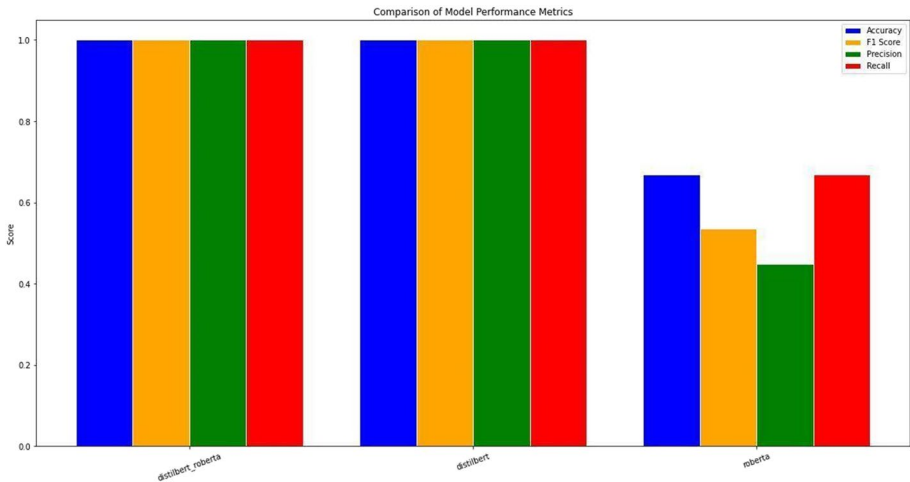


Fig. 14 Classification performance of DISTILBERT, ROBERTA models with DISTILBERT + ROBERTA

The classification performance of the DISTILBERT and ROBERTA models with combined DISTILBERT + ROBERTA model is shown in Table 12 and Figure 14.

#### Case 4: DISTILBERT, BERT with DISTILBERT + BERT

The classification performance of the DISTILBERT and BERT models with combined DISTILBERT + BERT model is shown in Table 13 and Figure 15.

#### Case 5: DISTILBERT, BERT with DISTILBERT + BERT

The classification performance of ROBERTA and BERT models with the ROBERTA + BERT model is shown in Table 14 and Figure 16.

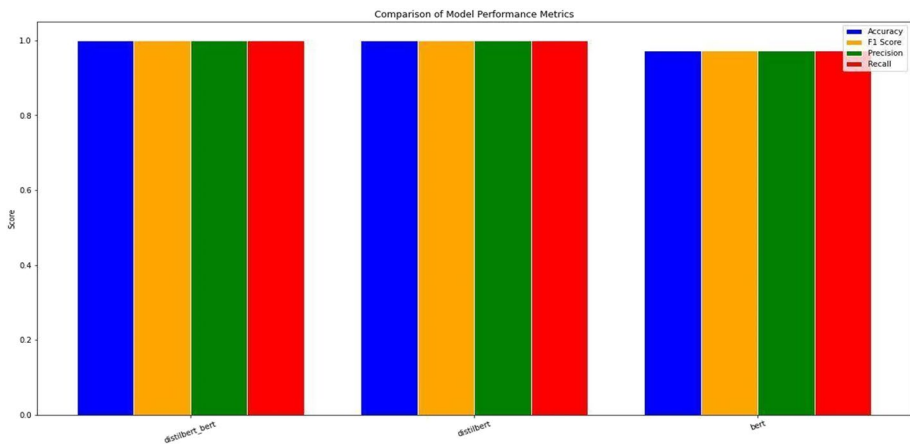
#### Case 6: Combined dual models

The classification performance of the combined dual models is shown in Table 15 and Figure 17.

As inferred from the classification report, the order of the ABSA performance of different models can be represented in the decreasing order as ROBERTA, BERT, and DISTILBERT models with a classification accuracy of 67 %, 97 %, and 100 % respectively. In

**Table 13** Classification report of DISTILBERT, BERT models with DISTILBERT + BERT

Model	Accuracy	F1 score	Precision	Recall
DISTILBERT + BERT	100%	100%	100%	100%
DISTILBERT	100%	100%	100%	100%
BERT	97.24%	97.19%	97.35%	97.24%



**Fig. 15** Classification performance of DISTILBERT, BERT models with DISTILBERT + BERT

**Table 14** Classification report of ROBERTA, BERT models with ROBERTA + BERT

Model	Accuracy	F1 score	Precision	Recall
ROBERTA + BERT	66.89%	53.62%	44.75%	66.89%
ROBERTA	66.89%	53.62%	44.75%	66.89%
BERT	97.24%	97.19%	97.35%	97.24%

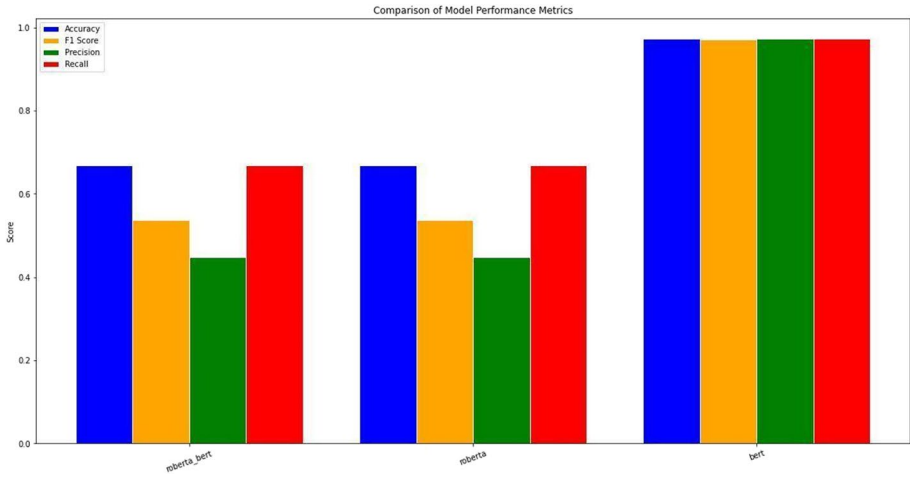


Fig. 16 Classification performance of ROBERTA, BERT models with ROBERTA + BERT

Table 15 Combined dual models classification report

Model	Accuracy	F1 score	Precision	Recall
DISTILBERT+R OBERTA	100%	100%	100%	100%
DISTILBERT+ B ERT	100%	100%	100%	100%
ROBERTA + BERT	66.89%	53.62%	44.75%	66.89%
Combined-dual model	100%	100%	100%	100%

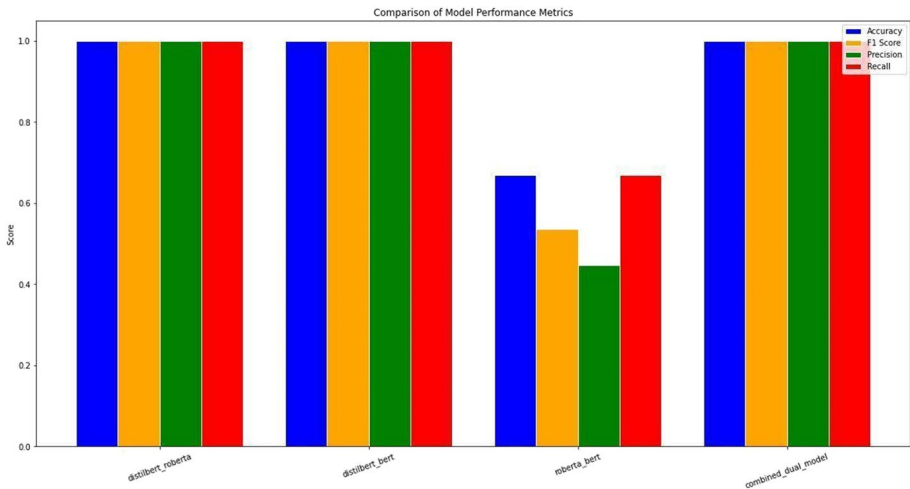


Fig. 17 Classification performance of combined dual models

all cases, the DISTILBERT model exhibits superior performance as both individual and hybrid classifier. Since all models were successful in identifying and classifying the sentiments from the data, it can be said that these models are effective in ABSA processes.

## 5.5 Discussion of Findings & Contribution

We have implemented a novel approach in this study related to the HyBERT model, which focuses on the strengths of transformer models like BERT, ROBERTA, and DISTILBERT and enhances aspect-based sentiment analysis (ABSA) and manages sentiment classification. Using textual data sourced from the Hugging Face Dataset, our research is centred on analysing the sentiments across different aspects. Our investigation encompasses a wide range of case analyses that assess the performance of ABSA on an individual and combined basis.

## 5.6 Analysis of Transformer Model Performance

Our evaluations were conducted by utilising the essential performance metrics that include the F1 score, recall, accuracy, and support. Undoubtedly, with individual applications or when employed with the BERT and RoBERTa models, the DISTILBERT model exhibits an impressive accuracy of 100%. This detailed categorization uncovers the consistency of this achievement by taking into consideration all combinations of all models. The BERT model, when closely observed, demonstrated a strong accuracy rate of 97%. Contrastingly, the performance of the ROBERTA model was lower and yielded an average accuracy of just 67%.

## 6 Validation Of HyBERT Model

The outcomes of our research validate the effectiveness of our proposed ByBERT model for real-time ABSA applications. Our model outperforms other models in terms of accuracy for both aspect-based sentiment analysis and sentiment categorization. These high levels of remarkable accuracy, which are achieved by the HyBERT model, reveal the potential to enhance the decision-making process based on reliable sentiment insights.

### 6.1 Significance of Future Implications

Our research has substantial applications in the fields of sentiment analysis and natural language processing. By blending the strengths of the transformer models, we extend a remarkable understanding of the sentiment complications represented in the textual data. Moving ahead, our study encourages the deployment of more difficult assemblies of transformer models that further elevate accuracy and applications. In addition to this, the principles and insights gained have also extended to other tasks that are text-related, which further contributes to deeper comprehensions of language-context interactions.

### 6.2 Real-time case study

These days, rating and reviewing culinary recipes, as well as submitting, looking for, and downloading them, have become daily routines. On YouTube, millions of users are looking

to exchange recipes. Through user feedback, a user spends a lot of time looking for the greatest cooking recipe. Sentiment analysis and opinion mining are essential tools for gathering information to ascertain what people are thinking. This sentiment-based real-time system that mines YouTube meta-data (likes, dislikes, views, and comments) in order to extract significant aspects related to culinary recipes and detect opinion polarity in accordance with these qualities is a real-time case study. To enhance the functionality of the system, a few authors worked on the vocabulary of cooking instructions and suggested some algorithms built on sentiment bags, depending on certain terms connected to food emotions and injections.

Or another example can be Aspect-Based Sentiment Analysis, which provides valuable insights into different aspects of the restaurant's performance, allowing the management to focus on improving specific areas to enhance the overall customer experience. For example, the restaurant might want to work on speeding up service and addressing pricing concerns to attract more customers and improve their overall rating.

### 6.3 Predominance of the Study Over Existing Studies of the Domain

The predominance of this study is largely influenced and encapsulated in the innovative approach, which involves a comprehensive evaluation of the transformer-based models for ABSA and the classification of sentiments. While previous studies have contributed valuable insights into sentiment analysis and ABSA, our study has a distinctive aspect that further elevates its significance.

**HyBERT Model:** Our study is an introduction to the HyBERT model, which is a novel hybrid approach that extends the power of the BERT, ROBERTA, and DISTILBERT transformer models. This integration harnesses the unique power of each model, which helps in achieving superior accuracy and mixed sentiment analysis.

**Comprehensive Performance Analysis:** Evaluation of the Performance of Transformer Models that Cover Multiple Aspects of ABSA and Sentiment Classification This study is a systematic comparison of individual models that combine models and other model combinations that leverage a comprehensive understanding of capabilities.

**Diversified Evaluation Metrics:** Employing evaluation metrics inclusive of F1 score, recall, accuracy, and support that comprehensively assess the performance of models The holistic approach is an assurance that our findings are robust and reliable and capture various performance dimensions.

**Future-Ready Recommendations:** Our study aims for results and makes actionable recommendations for future research and challenges in implementation. Highlighting complex avenues leverage strategies while we feel implied towards the advanced and impactful sentiment analysis techniques.

## 7 Conclusion

Our study not only introduces the concept of a novel ByBERT model that significantly improves ABSA performance but also uncovers the importance of transformer-based techniques through sentiment analysis. These particular strengths and synergies among the BERT, RoBERTa, and DistilBERT models are examples of enhanced accuracy and insight generation. Our research actually paves the way for more sophisticated and effective methodologies related to sentiment analysis, ultimately facilitating the process of more

informed decision-making across various domains. Furthermore, our research lays the path for accepting transformer-based approaches in the intricate ABA scenarios, which therefore usher in a more detailed understanding of sentiment with multiple textual domains.

A novel HybBERT model is presented in this paper, which combines the efficacy of the transformer models such as BERT, ROBERTA, and DISTILBERT for aspect-based sentiment analysis and sentiment classification. The models are trained with the textual data containing aspects from the Hugging Face dataset to analyze the sentiments of the textual data based on different aspects. This research conducted different case analysis for analyzing the performance of the transformer models, the performance of combined dual models, and the performance of different combinations of the models for ABSA. The proposed approach's performance was evaluated by simulating the gathered data with respect to several performance measures such as F1 score, recall, accuracy, and support. When used alone or in conjunction with the BERT and ROBERTA models, the DISTIBERT model's accuracy can reach 100%, as can be seen from the categorization report for all potential combinations. The BERT model, on the other hand, obtained the second-best result with an accuracy of 97%. The ROBERTA model underperforms compared to other models, with an average accuracy of 67%. The results verify the proposed HybBERT model's usefulness for performing real-time ABSA by attaining superior accuracy for both aspect-based sentiment analysis and sentiment categorization.

**Data availability** This is the dataset link which will be used in the code:

<https://huggingface.co/datasets/Yaxin/SemEval2014Task4Raw/viewer/All/train>

## Declarations

**Conflicts of interests** Anushree Goud, Bindu Garg.

Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune.

Date:19/10/2023.

Dear Multimedia Tools and Applications Journal,

We wish to submit an original research article entitled "A Novel Framework for Aspect Based Sentiment Analysis Using a Hybrid BERT (HybBERT) model" for consideration by Multimedia Tools and Applications Journal. We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

We believe that this manuscript is appropriate for publication because It is feasible to extract the sentiment from the data by examining its numerous elements using an aspect-based sentiment analysis text analysis technique known as (ABSA). The ABSA approach analyses textual information and links the emotions to various textual elements. Due to its effectiveness, ABSA has recently grown significantly in importance. The researchers are looking at new methodologies and techniques for dealing with complex ABSA scenarios because of the growing importance of ABSA.

Please address all correspondence concerning this manuscript to me at [anushree.anushree.goud@gmail.com](mailto:anushree.anushree.goud@gmail.com).

Thank you for your consideration of this manuscript.

Sincerely,

Anushree Goud.

Bindu Garg.

## References

1. Mercha EM, Benbrahim H (2023) Machine Learning and Deep Learning for sentiment analysis across languages: A survey. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2023.02.015>
2. Chandra Y, Jana A (2020) Sentiment Analysis using Machine Learning and Deep Learning, 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), <https://doi.org/10.23919/indiacom49435.2020.9083703>

3. Gadri S, Chabira S, Ould Mehieddine S, Herizi K (2022) Sentiment Analysis: Developing an Efficient Model Based on Machine Learning and Deep Learning Approaches, *Intell Comput Optim*, pp. 237–247, [https://doi.org/10.1007/978-3-030-93247-3\\_24](https://doi.org/10.1007/978-3-030-93247-3_24)
4. Liu H, Chatterjee I, Zhou M, Lu XS, Abusorrah A (2020) Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods. *IEEE Trans Comput Soc Syst* 7(6):1358–1375. <https://doi.org/10.1109/tcss.2020.3033302>
5. Do HH, Prasad P, Maag A, Alsadoon A (2019) Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Exp Syst Appl* 118:272–299. <https://doi.org/10.1016/j.eswa.2018.10.003>
6. Rezaeinia SM, Rahmani R, Ghodsi A, Veisi H (2019) Sentiment analysis based on improved pre-trained word embeddings. *Expert Syst Appl* 117:139–147. <https://doi.org/10.1016/j.eswa.2018.08.044>
7. Liao S, Wang J, Yu R, Sato K, Cheng Z (2017) CNN for situations understanding based on sentiment analysis of twitter data. *Procedia Comput Sci* 111:376–381. <https://doi.org/10.1016/j.procs.2017.06.037>
8. Usama M, Ahmad B, Song E, Hossain MS, Alrashoud M, Muhammad G (2020) Attention-based sentiment analysis using convolutional and recurrent neural network. *Futur Gener Comput Syst* 113:571–578. <https://doi.org/10.1016/j.future.2020.07.022>
9. Gandhi UD, Malarvizhi Kumar P, Chandra Babu G Karthick G (2021) Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM), *Wirel Person Commun*, <https://doi.org/10.1007/s11277-021-08580-3>
10. Song Wei, Wen Zijian, Xiao Zhiyong, Park Soon Cheol (2021) Semantics perception and refinement network for aspect-based sentiment analysis. *Knowl-Based Syst* 214(28):106755. <https://doi.org/10.1016/j.knosys.2021.106755>
11. Wu J-L, He Y, Yu L-C, Lai KR (2020) Identifying Emotion Labels From Psychiatric Social Texts Using a Bi-Directional LSTM-CNN Model. *IEEE Access* 8:66638–66646. <https://doi.org/10.1109/access.2020.2985228>
12. Aishwarya R, Ashwatha C, Deepthi A, Beschi Raja J (2019) A Novel Adaptable Approach for Sentiment Analysis, *Int J Sci Res Comput Sci, Eng Inf Technol*, pp. 254–263, <https://doi.org/10.32628/cseit195263>
13. Akhtar MS, Chauhan DS, Ghosal D, Poria S, Ekbal A, Bhattacharyya P (2019) Multi-task learning for multi-modal emotion recognition and sentiment analysis. *arXiv preprint arXiv:1905.05812*, <https://doi.org/10.48550/arXiv.1905.05812>
14. Bao L, P Lambert, Badia T (2019) Attention and Lexicon Regularized LSTM for Aspect-based Sentiment Analysis,” *ACLWeb*. <https://aclanthology.org/P19-2035/> (accessed Dec. 16, 2022)
15. Xing B, Liao L, Song D, Wang J, Zhang F, Wang Z, Huang H (2019) Earlier attention? aspect-aware LSTM for aspect-based sentiment analysis. *arXiv preprint arXiv:1905.07719*. <https://doi.org/10.48550/arXiv.1905.07719>
16. Xu B, Wang X, Yang B, Kang Z (2020) Target Embedding and Position Attention with LSTM for Aspect Based Sentiment Analysis, *Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence*, <https://doi.org/10.1145/3395260.3395280>
17. Al-Smadi M, Talafha B, Al-Ayyoub M, Jararweh Y (2018) Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *Int J Mach Learn Cybern* 10(8):2163–2175. <https://doi.org/10.1007/s13042-018-0799-4>
18. Alexandridis G, Aliprantis J, Michalakos K, Korovesis K, Tsantilas P, Caridakis G (2021) A Knowledge-Based Deep Learning Architecture for Aspect-Based Sentiment Analysis. *Int J Neural Syst* 31(10):2150046. <https://doi.org/10.1142/s0129065721500465>
19. Sivakumar M, Uyyala SR (2021) Aspect-based sentiment analysis of mobile phone reviews using LSTM and fuzzy logic. *Int J Data Sci Anal*. <https://doi.org/10.1007/s41060-021-00277-x>
20. Zhang K, Zhang K, Zhang M, Zhao H, Liu Q, Wu W, Chen E (2022) Incorporating dynamic semantics into pre-trained language model for aspect-based sentiment analysis. *arXiv preprint arXiv:2203.16369*, <https://doi.org/10.48550/arXiv.2203.16369>
21. Shim H, Lowet D, Luca S, Vanrumste B (2021) LETS: A Label-Efficient Training Scheme for Aspect-Based Sentiment Analysis by Using a Pre-Trained Language Model. *IEEE Access* 9:115563–115578. <https://doi.org/10.1109/access.2021.3101867>
22. Troya A, Gopalakrishna Pillai R, Rodriguez Rivero C, Genç Z, Kayal S, Araci D (2021) Aspect-Based Sentiment Analysis of Social Media Data With Pre-Trained Language Models, *2021 5th International Conference on Natural Language Processing and Information Retrieval (NLPIR)*, <https://doi.org/10.1145/3508230.3508232>
23. Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding *arXiv preprint arXiv:1810.04805*, <https://doi.org/10.48550/arXiv.1810.04805>

24. Wang W, Bi B, Yan M, Wu C, Bao Z, Xia J, Si L (2019) Structbert: Incorporating language structures into pre-training for deep language understanding arXiv preprint arXiv:1908.04577, <https://doi.org/10.48550/arXiv.1908.04577>
25. Batra H, Punn NS, Sonbhadra SK, Agarwal S (2021) BERT-Based Sentiment Analysis: A Software Engineering Perspective, *Lect Notes Comput Sci*, pp. 138–148, [https://doi.org/10.1007/978-3-030-86472-9\\_13](https://doi.org/10.1007/978-3-030-86472-9_13)
26. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le QV (2019) XLNet: Generalized Autoregressive Pretraining for Language Understanding arXiv preprint arXiv:1906.08237, <https://doi.org/10.48550/arXiv.1906.08237>. Working
27. Mao Y, Wang Y, Wu C, Zhang C, Wang Y, Yang YY, Bai J (2020) Ladabert: Lightweight adaptation of bert through hybrid model compression”, arXiv preprint arXiv:2004.04124, <https://doi.org/10.48550/arXiv.2004.04124>
28. Sun C, Huang L, Qiu X (2019) Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence, arXiv preprint arXiv:1903.09588, <https://doi.org/10.48550/arXiv.1903.09588>
29. Li X, Bing L, Zhang W, Lam W (2019) Exploiting BERT for end-to-end aspect-based sentiment analysis, arXiv preprint arXiv:1910.00883, <https://doi.org/10.48550/arXiv.1910.00883>
30. Adoma AF, Henry N-M, Chen W (2020) Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition, 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), <https://doi.org/10.1109/iccwamtip51612.2020.9317379>
31. Phan MH, Ogunbona P O (2020) Modelling Context and Syntactical Features for Aspect-based Sentiment Analysis, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/2020.acl-main.293>
32. van den Broek-Alteneburg EM, Atherly AJ (2019) Using Social Media to Identify Consumers’ Sentiments towards Attributes of Health Insurance During Enrollment Season, *Appl Sci MDPI*, <https://doi.org/10.3390/app9102035>
33. Meng W, Wei Y, Liu P, Zhu Z, Yin H (2019) Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM. *IEEE Access* 7:167240–167249. <https://doi.org/10.1109/access.2019.2952888>
34. Zhang W, Li X, Deng Y, Bing L, Lam W (2022) A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges, *IEEE Trans Knowl Data Eng*, pp. 1–20, <https://doi.org/10.1109/tkde.2022.3230975>
35. Howard J, Ruder S (2018) Universal language model fine-tuning for text classification, arXiv preprint arXiv:1801.06146, <https://doi.org/10.48550/arXiv.1801.06146>
36. Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training, [https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf)
37. Azhar AN, Khodra ML (2020) Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis, 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA), <https://doi.org/10.1109/icaicta49861.2020.9428882>
38. Bhuvaneshwari P, Rao AN, Robinson YH et al (2022) Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model. *Multimed Tools Appl* 81:12405–12419. <https://doi.org/10.1007/s11042-022-12410-4>
39. Sanh V, Debut L, Chaumond J, Wolf T (2019) DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, arXiv preprint arXiv:1910.01108, <https://doi.org/10.48550/arXiv.1910.01108>
40. Ruz GA, Henríquez PA, Mascareño A (2020) Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Futur Gener Comput Syst* 106:92–104. <https://doi.org/10.1016/j.future.2020.01.005>
41. Li Z, Fan Y, Jiang B et al (2019) A survey on sentiment analysis and opinion mining for social multimedia. *Multimed Tools Appl* 78:6939–6967. <https://doi.org/10.1007/s11042-018-6445-z>
42. Geetha MP, KarthikaRenuka D (2021) Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *Int J Intell Netw* 2:64–69. <https://doi.org/10.1016/j.ijin.2021.06.005>
43. Li Hengyun, Yu Bruce X.B., Li Gang, Gao Huicai (2023) Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews. *Tourism Manag* 96:104707. <https://doi.org/10.1016/j.tourman.2022.104707>
44. Gupta H, Goyal SS, Sawant Swaroop A Baral MC (2019) Target-oriented Opinion Words Extraction with Target-fused Neural Sequence Labeling, <https://arxiv.org/pdf/2109.08079.pdf>



45. Xing B, Liao L, Song D, Wang J, Zhang F, Wang Z, Huang H (2019) Earlier attention? aspect-aware LSTM for aspect-based sentiment analysis, arXiv preprint arXiv:1905.07719, <https://doi.org/10.48550/arXiv.1905.07719>
46. Xu L, Bing L, Lu W, Huang F (2020) Aspect Sentiment Classification with Aspect-Specific Opinion Spans, <https://aclanthology.org/2020.emnlp-main.288.pdf>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.