Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2023 SCPE. Volume 24, Issues 3, pp. 277–286, DOI 10.12694/scpe.v24i3.2444

IMPROVING BERT MODEL ACCURACY FOR UNI-MODAL ASPECT-BASED SENTIMENT ANALYSIS TASK

AMIT CHAUHAN AND RAJNI MOHANA*

Abstract. Techniques and methods for examining users' feelings, emotions, and views in text or other media are known as "sentiment analysis," this phrase is used frequently. In many areas, including marketing and online social media, analysis of user and consumer opinions has always been essential to decision-making processes. The development of new methodologies that concentrate on analysing the sentiment associated with specific product characteristics, such as aspect-based sentiment analysis (ABSA), was prompted by the need for a deeper understanding of these opinions. Despite the growing interest in this field, some misunderstanding exists about ABSA's core ideas. Even though sentiment, affect, emotion, and opinion refer to various ideas, they are frequently used synonymously. This ambiguity commonly causes user opinions to be analysed incorrectly. This work provides an overview of ABSA and the issue of overfitting. Following this analysis, we improved the model by enhancing the accuracy and F1 score of the existing model by fine-tuning the technique. Our model outperformed the others, achieving the best results for the restaurant dataset with an 85.02 accuracy and a 79.19 F1 score, respectively.

Key words: Aspect-based sentiment analysis, Sentiment analysis, Natural language processing, Online Social networks, Opinion Mining.

1. Introduction. ABSA is a technique used in Natural Language Processing (NLP) to ascertain a text's attitude towards a particular quality or attribute of a good, service, or institution. It entails determining the sentiment polarity (neutral, positive or negative) for a specific interest feature, such as a smartphone's battery life or a cuisine's flavour. This type of analysis is essential in several industries, including marketing, product development, and customer service, as it provides valuable insights into customer opinions and preferences.ABSA frequently includes NLP techniques, such as sentiment analysis (SA), named object identification, and topic modelling [27].

Meanwhile, unimodal SA analyses unimodal data, such as text from [25], to determine people's attitudes towards various topics or products. Online platforms like Twitter and Facebook have become well-liked methods to share and post views, including text, pictures, and audio, as social media and web technology have grown in popularity.

Analysis of the sentiment of a sentence towards particular opinion objectives is the job involved in ABSA. This includes identifying the polarity of sentiment (positive, neutral or negative) for each target mentioned in the statement, such as "SERVICE" and "FOOD," which would be negative and positive, respectively [24].

For example, consider a real-life scenario where you work for an e-commerce company that analyses customer reviews of a new smartphone. Your task is to perform ABSA on the reviews to obtain information about the aspects customers discuss and their sentiments towards each element. In the given study, "I like the design of the phone, but the battery life is a major disappointment," there are two aspects - A1: design and A2: battery life. Design is a positive aspect, while battery life is a negative aspect.

Previous studies have primarily relied on RNNs to model text, which suffers from complex parallelization, memory and computation requirements, and the inability to capture long-term dependencies due to truncated backpropagation through time (BPTT). Although LSTMs can maintain long-range information and partially solve the issue of disappearing gradients, they need a lot of training data. The label unreliability problem, particularly for neutral sentiment, a fuzzy condition that makes learning more difficult, is another difficulty that has been ignored in earlier studies.

*Jaypee University of Information Technology, Waknaghat Solan, Himachal Pradesh, India (chauhanamit37@gmail.com, rajni.mohana@juitsolan.in)

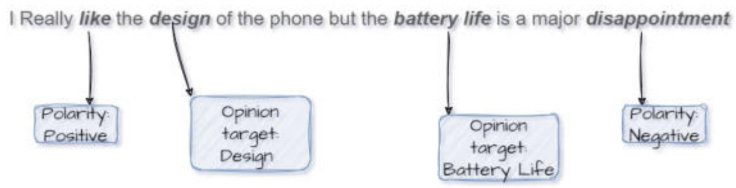

Fig. 1.1: Exemplary ABSA Statement

Targeted sentiment classification involves identifying sentiment towards a specific target entity or aspect in text, but label unreliability is a common challenge in this task [25]. This refers to the inconsistency or inaccuracy of the labelled data used to train a sentiment classifier due to subjective or ambiguous sentiment labels assigned by different annotators. One approach uses inter-annotator agreement metrics to address this issue and filter out instances where annotators disagree. In contrast, another uses active learning to select informative or uncertain illustrations for annotation and update the model iteratively. Semi-supervised learning can also leverage improving both labelled and unlabeled data classifier performance and robustness.

To enhance the effectiveness of the essential BERT model for sentiment classification, especially for fuzzy labels, this research suggests a model that uses label smoothing regularisation [24]. Our model outperforms basic BERT when used with pre-trained BERT on three benchmark datasets, demonstrating competitive performance and lightweight properties. The paper's key contributions include addressing overfitting issues, applying pre-trained BERT, and using fine-tuning modified BERT SPC model to overcome overfitting in target and context words.

1.1. ***Motivation.*** This study's primary objective is to assess sentiment in opinionated text where users provide feedback about products, services, or events through writing reviews. By defining terms like emotion and affective opinions, an effective ABSA model must be able to recognise different sentimental dimensions. The literature review section highlights the need for a commonly agreed-upon definition of fundamental ABSA concepts and appropriate metrics for each measurement. Positivity about a product differs from expressing sentiment towards it and thus demands various measurement and evaluation methods. While computer scientists may conflate feelings and opinions, this is not true in other fields such as social sciences, psychology, marketing and communications, and neuroscience. This divergence poses a genuine obstacle to creating reliable and practical real-world applications. Defining the properties of the document or media being analysed is a crucial issue that ABSA models must handle. Review analysis primarily concentrates on user-generated text. However, contemporary websites and social media platforms provide several tools and features that let consumers write in-depth and elaborate reviews of goods or services. In addition to the main text, users can provide ratings on a five-star scale, add tags to categorize their thoughts, use emoticons to convey their sentiments, and more. Traditional ABSA techniques typically consider only some of these elements.

2. Related Work. ABSA is a powerful NLP technique that indicates a sentence's polarity for a specific aspect term. Giuseppe D'Aniello et al. (2022) [8] have provided an overview of the latest ABSA methodologies and approaches, highlighting the significant challenges of recent developments in this area. Their research introduces a new reference model called KnowMIS-ABSA, which can be extended to include reviews and other ABSA features in future work. Bin Liang et al. (2022) [9] introduced Sentic GCN, a graph convolutional network that utilizes emotional relationships of text for a specific aspect by using SenticNet. They use dynamic knowledge to enhance sentence dependency graphs and build unique chart neural networks. Haiyan Wu et al. (2022) [20] have introduced an attention network with phrase dependency for the ABSA task (PD-RGAT). The phrase dependency graph is created using this relational graph attention network by combining directed dependency edges and phrase information. GloVe and BERT, two distinct pre-training models, were tested and obtained results comparable to other baseline models. Kai He et al. (2022) [14] the meta-based MSM self-training system, which has a meta-weighter, was proposed by the authors. According to the authors, a neural system with effective learning control and appropriate symbolic representation choice can produce a generalizable model. They use MSM to train a teacher model to create in-domain knowledge; a student model subsequently uses it for supervised learning. Ziguo Zhao et al. (2022) [26] a graph convolutional network with several weighting strategies was developed for aspect-based sentiment analysis. To fully utilize BERT, they introduced a dynamic weight alignment technique. They also devised an aspect-aware weight mechanism to govern message propagation to the aspect. They introduced an aspect-oriented loading layer to limit the negative impacts of words unrelated to the aspect term. Finally, they fused high-order semantic and grammatical information via multi-head self-attention to forecast premium aspect-specific representations. Li Yang et al. (2022) [22]developed the Cross-Modal Multitask Transformer (CMMT), a multitask learning framework that includes two additional activities to practise aspect- and emotion-aware intra-modal examples. They also present a Text-Guided Cross-Modal Communication Module to dynamically regulate the contributions of visual information to each word's representation in the inter-modal interaction. Shi Feng et al. (2022) [11]proposed a new ABSA model called AG-VSR that relies on aspect representations rather than sentence representations and updates them with GCNs. To overcome the dependency on the integrity of the dependency tree and retain global sentence information, they used Attention-assisted Graph-based Representation (A GR) and Variational Sentence Representation (VSR). The GCN module generates a GR by modifying a dependency tree with the attention mechanism. Lastly, [4] Anan Dai et al. (2022) proposed a human cognition-based approach to It entails instruction in everything from sentence grammar to word meaning. To capture words' structural and broad semantics, they created a dual-channel semantic learning graph convolutional network (GCN). They then carried out a syntactic GCN to learn the syntactic structure of sentences. Their approach aligns with human processing practices and provides a meaningful interpretation of the given sentence.

3. Sentiment Analysis. Opinion Mining (OM) and Sentiment Analysis have emerged as established research fields over the last twenty years, finding widespread use in various applications, including commercial ones. The definitions of OM and SA and any distinctions between opinion, sentiment, emotion, affect, and related concepts continue to be unclear despite the enormous advancement and innovation in these disciplines [7]. Many academics contend that these notions are separate and call for various methodologies that result in multiple conclusions, even though some see this as a dispute over language with no genuine difference between these concepts. Despite the fact that both OM and SA seek to grasp the subjectivity expressed in a text, the exact definitions of opinion and sentiment are up for debate among scholars.

Sentiment analysis involves analyzing opinions at various levels, including the sentence, document, aspect, and concept levels. The Document-level analysis aims to learn the polarity of an entire document, such as a review or news article, using a single positive or negative score to condense the contradiction of several words. Approaches based on lexicons are frequently employed in document-level analysis [2], determining the polarity of each term before combining them to categorise the overall sentiment of the material. However, some researchers argue that considering the importance of each word could improve classification accuracy [3].

Sentence-level SA aims to understand a single sentence's negative or positive opinion. This method is frequently used to evaluate the sentiment of an entire document by combining the polarity scores of each sentence. While sentiment analysis at the text or sentence level offers helpful insights into users' perceptions of particular entities, it does not reveal the specific elements, features, or aspects that influence user opinions. Aspect-level sentiment analysis is required for this type of analysis. This approach involves identifying each aspect's opinionated sentences, entity categories, elements, linguistic expressions, and polarities. Concept-level SA, on the other hand, uses semantic analysis methods to find and examine concepts in the text, allowing computers to comprehend natural language and emotions at a deep level. With encouraging outcomes, SenticNet is a lexical resource that assigns mood and emotion tags to facilitate concept-level SA.

3.1. ***Aspect-Based Sentiment Analysis (ABSA).*** The goal of ABSA, a technique for natural language processing, is to extract sentiment information from text based on various characteristics or attributes of a given subject or entity. Sentiment identification, the next stage in ABSA, requires calculating a sentiment score (regarding orientation or polarity) for every aspect obtained from the text. The score typically falls within a mathematical range, such as a five-point scale or a decimal number between [-1, +1], where -1 was highly damaging, and +1 was highly positive. It expresses the degree of positivity or negativity regarding the element. The task is also known as sentiment classification if the only objective is to ascertain whether the sentiment is favourable or adverse [19]. A thorough analysis of contemporary aspect extraction methods is conducted, and they are divided into three primary categories: unsupervised, semi-supervised, and supervised procedures [23].

The unsupervised category includes techniques such as frequency or statistics-based heuristics, syntactic dependency path-based approaches, and rule-based techniques. The semi-supervised category comprises lexicon-based, dependency tree-based, and graph-based, for example, methods. Supervised techniques, on the other hand, employ machine learning approaches such as random fields, SVM, decision trees, autoencoders and neural networks [13]. Many methods for identifying sentiment can be categorised into three groups: lexicon-based, machine learning-based, and hybrid methods [21]. Let me explain ABSA with a case study from real life.

Consider a client who wants to reserve a hotel for her future vacation. She begins her study online and pursues numerous evaluations of hotels in a specific tourist area. She discovers a hotel review: ”Great amenities and location; however, the service was lacking. The staff was impolite and uncooperative. Although the breakfast was not up to standard, the hotel was clean and big. The location, facilities, service, staff, room, and breakfast are a few of the numerous characteristics listed in the review that may be analysed using ABSA in this case. Let’s see how ABSA might apply here:

Location: The reviewer described it as ”great”; the tone is favourable.

Amenities: The critic did not identify any problems. Thus, the sentiment is favourable.

Service: The reviewer expressed dissatisfaction with the service, calling it ”rude,” ”disappointing,” and ”unhelpful.”

Employees: The reviewer’s comments about the employees being ”rude” and ”unhelpful” reflect the overall poor opinion.

I had a positive feeling because it was described as ”clean” and ”spacious” in the room evaluation.

Breakfast: The reviewer expressed dissatisfaction, calling it ”subpar”.

To help her comprehend that while the location and amenities are excellent, the service and breakfast components are unfavourable, ABSA assists in extracting sentiment information from various aspects of the evaluation. Based on the most critical factors to the client, this information can help her make an informed decision about whether or not to reserve the hotel.

4. The Proposed Method. The study’s authors utilized the SemEval 2014 Task 4 Public dataset for data collection in the first step. Figure 2 depicts the flow of the proposed research, where the second step involves data pre-processing. The following step is feature selection, followed by training until the early stop prompt is received. The K-fold cross-validation (K=6) is then implemented in stage 2. If the accuracy is unsatisfactory, hyperparameter tuning is applied to the model for better results. The pre-trained models of the embedding layer played a vital role in achieving better results [18].

The embedding layer is a crucial component of neural network models in NLP. It converts textual data, such as words or sentences, into numerical vectors that the model can process [5]. The embedding layer maps each word or token in the input text to a high-dimensional embedding vector. The size of the embedding vector is typically smaller than the vocabulary size, making processing more efficient. Pre-trained embeddings trained on large datasets using techniques such as Word2Vec or Because GloVe captures the semantic and syntactic links between words it can be helpful in NLP tasks. During training, the embedding layer adjusts the embedding vectors based on the error calculated by subsequent neural network layers [16]. This enables the model to learn how to produce embedding vectors that faithfully reflect the task’s incoming data. Using an embedding layer has dramatically improved the accuracy of text classification, sentiment analysis [1], machine translation, and other NLP tasks by allowing the model to process textual data and capture complex word relationships efficiently.

The BERT (Bidirectional Encoder Representations from Transformers) model is a pre-trained language model that generates contextualized word embeddings for various NLP tasks. The mathematical procedure for computing embeddings uses input tokens, a pre-trained language model with L transformer layers, and a hidden size H in the BERT embedding [12]. The BERT model function takes an input sequence of tokens S and outputs a series of concealed states.

$$\mathrm{H} = \{\mathrm{h}_1,\ \mathrm{h}_2, \ldots, \mathrm{h}_\mathrm{n}\}\,, \tag{4.1}$$

Our research utilized Bert Embedding throughout the experimentation phase. We extracted the final hidden state corresponding to that token to acquire the BERT embedding for a particular permit or sub-word,

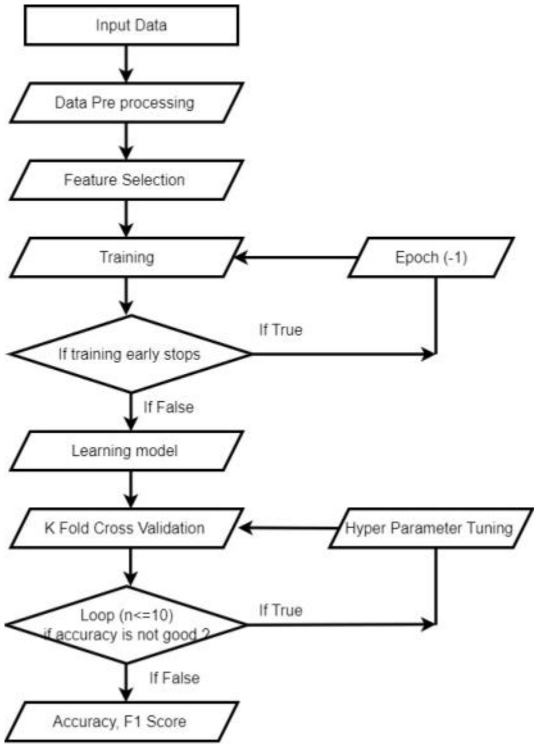

Fig. 4.1: The proposed Work

as outlined in equations 4.1 and 4.2. This remote state results after the input sequence passes through the L-th transformer layer. As a result, the BERT embedding of the i-th token can be expressed as equation 4.2.

$$\mathrm{BERT}(S)_i = h_i \wedge (L) \tag{4.2}$$

An alternative method of computing the BERT embedding involves computing the average or maximum hidden states of a sequence of tokens. Here, the set of indices I corresponds to the token sequence in S, and the BERT embedding of the sequence can be defined as equation 4.3.

$$\mathrm{BERT}(S)_I = \mathrm{mean}\,/\,\max\left(\{h_i \wedge (L) \mid i \text{ in } I\}\right) \tag{4.3}$$

To compute the BERT embedding, one can either take the final hidden state of a particular token or sub-word or average/take the maximum of the hidden forms of a sequence of tokens, as generated by the pre-trained BERT language model. The averaging or maximum operation is performed over the set of hidden states corresponding to the tickets in I.

5. Experimental Setup. Our experimentation used the Intel (R) Xeon (R) CPU E3-1225v5@3.31GHz Processor, 32 GB RAM, and a 64-bit-based processor.

5.1. *Dataset.* The SemEval-2014 Task 4 Subtask 2 dataset is intended for academics and programmers working on sentiment analysis and natural language processing tasks. The collection benefits academics who are building and testing algorithms for classifying the emotions of brief messages, like tweets. Researchers should utilise this data set because it has been annotated by qualified individuals with sentiment polarity tags. This indicates that the dataset offers robust ground truth data necessary for creating and testing precise sentiment analysis algorithms. The main reason for choosing the SemEval 2014 Task 4 subtask 2 dataset is because the set of data is a popular baseline data for evaluating sentiment analysis algorithms so that authors can compare

Table 5.1: Aspect words classified into positive, negative, and neutral polarities are present in all three datasets' train/test sets.

Dataset		**Instances**	**Positive**	**Negative**	**Neutral**
LAPTOP 14	Train	2,328	994	870	464
	Test	638	341	128	169
RESTAURANT 14	Test	1,106	728	182	196
	Train	3,608	2,164	807	637
Twitter 14	Train	6,248	1561	1,560	3,127
	Test	692	173	173	346

the work to a substantial body of existing research and a standard assessment process. In general, using the SemEval-2014 Task 4 Subtask 2 dataset can assist in creating and assessing cutting-edge sentiment analysis algorithms, comparing the results to an accepted standard, and contributing to the larger natural language processing research tasks.

To Perform our experimentation work, we employed the SemEval 2014 Task 4 Subtask 2 dataset [18]. The dataset comprises three domain datasets: Laptop, Twitter, and Restaurant. Each dataset contains separate train and test files.

This study includes experimentation using both a training testing set and a training training set to demonstrate that overfitting is a concern with the SemEval 2014 Task 4 dataset. When a higher epoch value is set, the early stop occurs during the training stage.

The number of times a model is trained on the full dataset is called an "epoch" during the training process [15]. Increasing the epoch value may allow the model to learn more complex patterns within the data. Still, it can also result in overfitting, where the model becomes too specialized in the training data, leading to poor performance on new data.

Early stopping methods are used during training to monitor a validation measure like accuracy or loss to improve the model's performance and prevent overfitting. The activity is terminated if the validation measure subsequently ceases to improve. In some circumstances, setting the highest epoch value may be suitable. Still, it's crucial to consider dataset size and model complexity when choosing the correct number of epochs. Early halting strategies that depend on validation measures work better in general.

To combat overfitting, we used the K-fold cross-validation approach. This method involves dividing the data into K similarly sized "folds," training the model on K-1 folds, and then testing it on the residual fold [10]. The process is repeated K times to get an overall assessment of the model's performance, with each fold serving as the test set once.

Cross-validation can provide a more accurate estimate of the model's performance than utilising a single train-test split to prevent overfitting by ensuring that the model is tested on a variety of data [28]. There are various types of k-fold cross-validation, such as stratified k-fold, which preserves the class distribution in each fold. Leave-one-out cross-validation is another variant, where each instance is used as the test set once, and K is set to the number of instances in the dataset.

Overall, K-fold cross-validation helps evaluate a model's performance and can identify potential issues such as overfitting. In mathematical notation [6, 17], the formula for K-fold cross-validation can be expressed as follows in equation 5.1: Let $X = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$ be a dataset of size N, and K be the number of folds. For each fold i in K, let

$$x_i = \left\{\left(x_{i\_1}, y_{i\_1}\right), \left(x_{i_2}, y_{i_2}\right), \ldots, \left(x_{i_{N_i}}, y_{i_{N_i}}\right)\right\} \tag{5.1}$$

be the i-th fold of size N_i, such that $N = k \cdot N_i$. Let

$$x_{\text{train}} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} - X_i \tag{5.2}$$

The (X train) dataset was used to train the model, and the parameter vector theta was obtained to evaluate X test and get the predicted values (y pred). The evaluation metric(s) used, such as mean squared error or

Table 5.2: Without the Cross-Validation Technique

Dataset	Training_Testing		Training_Training	
	Accuracy	F1	Accuracy	F1
Laptop	0.8058	0.7569	0.8990	0.8885
Twitter	0.7356	0.7011	0.9190	0.9180
Restaurant	0.8502	0.7919	0.9475	0.9224

Table 5.3: With K Fold Cross Validation Technique (K=6)

Dataset	Training_Testing		Training_Training	
	Accuracy	**F1**	**Accuracy**	**F1**
Laptop	0.8058	0.7569	0.8990	0.8885
Twitter	0.7356	0.7011	0.9190	0.9180
Restaurant	0.8502	0.7919	0.9475	0.9224

accuracy, were calculated between (y pred) and the actual values (y test) in (X test). This process was repeated for all i in K, using equation 5.2 to calculate the evaluation metric(s) each time.

The K-fold cross-validation technique was used to test the BERT-SPC model on three datasets, and the results are presented in Table 5.2. The restaurant dataset outperformed the other two datasets. During training, the authors observed that the "Early Stop" prompt occurred in every epoch. To address this, they implemented the K Cross-Validation technique with K=6. The overall performance estimate was obtained by averaging the K evaluation metric(s) over all folds.

Table 5.3 displays the results obtained using the K Cross-Validation Technique with k=6, which led to improved outcomes. This technique addressed the issue of overfitting, and the K Cross-Validation technique performed better than other approaches.

5.2. ***Model Comparison.*** A comparison of the proposed model (BERTSPC with K fold cross-validation) and (baseline research work) is shown in Table 5.4. The authors improved the performance of the BERTSPC model with the hyperparameters tuning technique. The authors encountered the overfitting issue during experimental work. To address this issue, they implemented the K Fold cross-validation technique, which yielded the best results for the BERTSPC model. The authors also experimented with several cross-validation techniques, such as the Stratified Cross-validation technique, Leave One Out Cross-validation, and Group K Fold Validation, on the SemEval 2014 Dataset and achieved comparable results using the Leave One Out and Group K Fold techniques. However, the K Fold Cross-validation technique produced the best results in the authors' experimentation.

Figure 3,4 and 5 shows the Accuracy and F1 Score of various baseline models for the Twitter, Restaurant and Laptop Dataset, respectively. From Figure 3 for the Twitter dataset in the previous BERTSPC model, the accuracy is 73.55, and in our proposed model, we achieved 0.01 high precision compared to the previous model. From Figure 3, we can see ATAE-LSTM and IAN are not plotted because, in previous studies, these models were not tested on this dataset. From Figure 4 for the [27]Restaurant dataset in the previous BERTSPC model, the accuracy is 84.46, and in our proposed model, we achieved 0.56 high precision compared to the previous model. The laptop dataset's highest level of accuracy is shown in Figure 5. From Figure 5, the accuracy of the laptop dataset in the last BERTSPC model is 73.55, and in our proposed model, we achieved 1.59 high precision compared to the previous model.

6. Conclusion. This study focuses on absa, which utilizes Attention-based Encoders for Target and Context. We used a pre-trained BERT model for this task to achieve the best possible outcomes. Through the process of hyperparameter tuning, we were able to enhance the accuracy of the Bert Spc model. However, during experimentation, we identified an overfitting issue addressed by implementing the Cross-validation tech-

Table 5.4: Baseline model findings are taken from articles that have already been published, (-) means not available

Models	Twitter		Restaurant		Laptop	
	Acc.	F1	Acc.	F1	Acc.	F1
TD LSTM	0.7080	0.6900	0.7563	-	0.6813	-
ATAE LSTM	-	-	0.7720	-	0.6870	-
IAN	-	-	0.7860	-	0.7210	-
RAM	0.6936	0.6730	0.8023	0.7080	0.7449	0.7135
FEATURE-BASED SVM	0.6340	0.6330	0.8016	-	0.7049	-
REC-NN	0.6630	0.6590	-	-	-	-
MEMNET	0.6850	0.6691	0.7816	0.6583	0.7033	0.6409
AEN- GloVe w/o PCT	0.7066	0.6907	0.8017	0.7050	0.7272	0.6750
AEN-GloVe w/o MHA	0.7124	0.6953	0.7919	0.7028	0.7178	0.6650
AEN-GloVe w/o LSR	0.7080	0.6920	0.8000	0.7108	0.7288	0.6869
AEN- GloVe BiLSTM	0.7210	0.7042	0.7973	0.7037	0.7312	0.6980
AEN GloVe	0.7283	0.6981	0.8098	0.7214	0.7351	0.6904
BERT SPC	0.7355	0.7214	0.8446	0.7698	0.7899	0.7503
AEN BERT	0.7471	0.7313	0.8312	0.7376	0.7993	0.7631
BERT SPC with K Fold Cross Validation (Proposed)	**0.7356**	**0.7011**	**0.8502**	**0.7919**	**0.8058**	**0.7569**

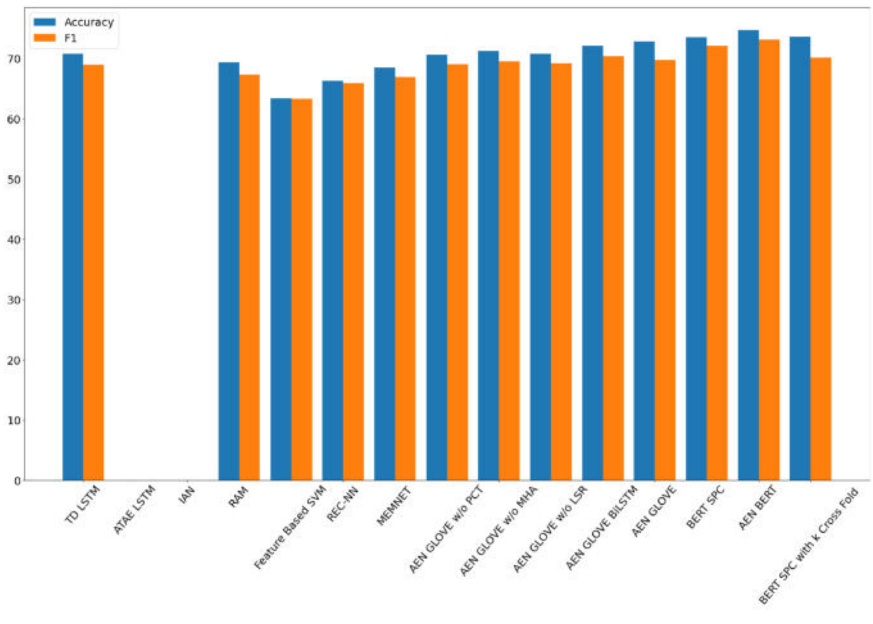

Fig. 5.1: Accuracy and F1 score on Twitter Data set

nique. Our findings indicate that the K Cross-validation approach with K=6 produced the best results. Moving forward, we plan to apply this approach to real-world datasets. While this is an unimodal problem, we intend to extend our policy to multimodal issues.

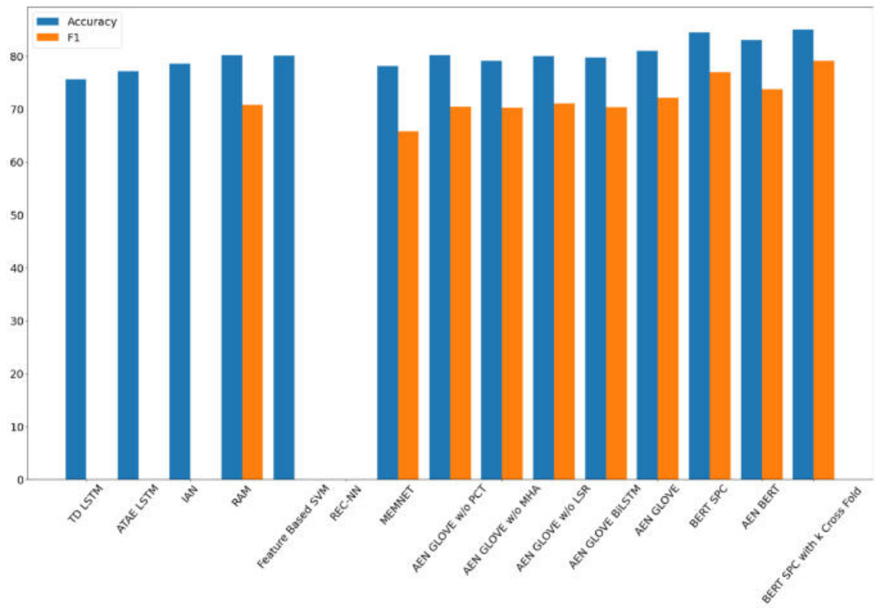

Fig. 5.2: Accuracy and F1 score on Restaurant Dataset

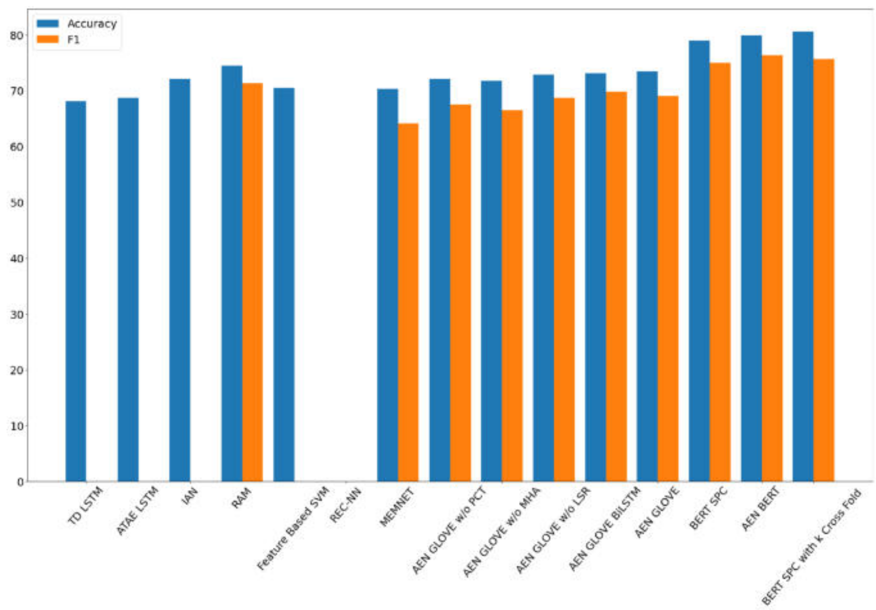

Fig. 5.3: Accuracy and F1 score on Laptop Dataset

REFERENCES

[1] W. T. Alshammari and S. AlHumoud, *Taqs: An arabic question similarity system using transfer learning of bert with bilstm*, IEEE Access, 10 (2022), pp. 91509–91523.

[2] A. Chauhan and R. Mohana, *Implementing lda topic modelling technique to study user reviews in tourism*, in 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC), IEEE, 2022, pp. 357–360.

[3] A. Chauhan and M. Sood, *Sybil account detection in online social networks using statistical feature selection techniques*, in

Futuristic Trends in Networks and Computing Technologies: Second International Conference, FTNCT 2019, Chandigarh, India, November 22–23, 2019, Revised Selected Papers 2, Springer, 2020, pp. 605–618.

[4] A. Dai, X. Hu, J. Nie, and J. Chen, *Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis*, International Journal of Data Science and Analytics, 14 (2022), pp. 17–26.

[5] Z. Ding, H. Li, W. Shang, and T.-H. P. Chen, *Can pre-trained code embeddings improve model performance? revisiting the use of code embeddings in software engineering tasks*, Empirical Software Engineering, 27 (2022), p. 63.

[6] J. D. Domingo, R. M. Aparicio, and L. M. G. Rodrigo, *Cross validation voting for improving cnn classification in grocery products*, IEEE Access, 10 (2022), pp. 20913–20925.

[7] G. Dubey, H. P. Singh, K. Sheoran, G. Dhand, and P. Malik, *Drug review sentimental analysis based on modular lexicon generation and a fusion of bidirectional threshold weighted mapping cnn-rnn*, Concurrency and Computation: Practice and Experience, (2023), p. e7512.

[8] G. D'Aniello, M. Gaeta, and I. La Rocca, *Knowmis-absa: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis*, Artificial Intelligence Review, 55 (2022), pp. 5543–5574.

[9] ———, *Knowmis-absa: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis*, Artificial Intelligence Review, 55 (2022), pp. 5543–5574.

[10] N. Elaraby, S. Barakat, and A. Rezk, *A conditional gan-based approach for enhancing transfer learning performance in few-shot hcr tasks*, Scientific Reports, 12 (2022), p. 16271.

[11] S. Feng, B. Wang, Z. Yang, and J. Ouyang, *Aspect-based sentiment analysis with attention-assisted graph and variational sentence representation*, Knowledge-Based Systems, 258 (2022), p. 109975.

[12] X. Feng and Y. Zeng, *Neural collaborative embedding from reviews for recommendation*, IEEE Access, 7 (2019), pp. 103263–103274.

[13] J. He, A. Wumaier, Z. Kadeer, W. Sun, X. Xin, and L. Zheng, *A local and global context focus multilingual learning model for aspect-based sentiment analysis*, IEEE Access, 10 (2022), pp. 84135–84146.

[14] K. He, R. Mao, T. Gong, C. Li, and E. Cambria, *Meta-based self-training and re-weighting for aspect-based sentiment analysis*, IEEE Transactions on Affective Computing, (2022).

[15] T.-h. Li, H. Zhang, L. Fan, H. Wang, and Q. Liu, *Research on deep neural network model construction and overfitting*, in International Conference on Neural Networks, Information, and Communication Engineering (NNICE), vol. 12258, SPIE, 2022, pp. 88–93.

[16] X. Li, X. Fu, G. Xu, Y. Yang, J. Wang, L. Jin, Q. Liu, and T. Xiang, *Enhancing bert representation with context-aware embedding for aspect-based sentiment analysis*, IEEE Access, 8 (2020), pp. 46868–46876.

[17] A. Maity, S. Ghosh, S. Karfa, M. Mukhopadhyay, S. Pal, and P. K. D. Pramanik, *Sentiment analysis from travellers' reviews using enhanced conjunction rule based approach for feature-specific evaluation of hotels*, Journal of Statistics and Management Systems, 23 (2020), pp. 983–997.

[18] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, *SemEval-2014 task 4: Aspect based sentiment analysis*, in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, Aug. 2014, Association for Computational Linguistics, pp. 27–35.

[19] T. A. Rana and Y.-N. Cheah, *Aspect extraction in sentiment analysis: comparative analysis and survey*, Artificial Intelligence Review, 46 (2016), pp. 459–483.

[20] H. Wu, Z. Zhang, S. Shi, Q. Wu, and H. Song, *Phrase dependency relational graph attention network for aspect-based sentiment analysis*, Knowledge-Based Systems, 236 (2022), p. 107736.

[21] K. Xu, F. Li, D. Xie, and D. Ji, *Revisiting aspect-sentiment-opinion triplet extraction: Detailed analyses towards a simple and effective span-based model*, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30 (2022), pp. 2918–2927.

[22] L. Yang, J.-C. Na, and J. Yu, *Cross-modal multitask transformer for end-to-end multimodal aspect-based sentiment analysis*, Information Processing & Management, 59 (2022), p. 103038.

[23] Q. Yang, Z. Kadeer, W. Gu, W. Sun, and A. Wumaier, *Affective knowledge augmented interactive graph convolutional network for chinese-oriented aspect-based sentiment analysis*, IEEE Access, 10 (2022), pp. 130686–130698.

[24] Q. Yong, C. Chen, Z. Wang, R. Xiao, and H. Tang, *Sgpt: Semantic graphs based pre-training for aspect-based sentiment analysis*, World Wide Web, (2023), pp. 1–14.

[25] G. Zhao, Y. Luo, Q. Chen, and X. Qian, *Aspect-based sentiment analysis via multitask learning for online reviews*, Knowledge-Based Systems, (2023), p. 110326.

[26] Z. Zhao, M. Tang, W. Tang, C. Wang, and X. Chen, *Graph convolutional network with multiple weight mechanisms for aspect-based sentiment analysis*, Neurocomputing, 500 (2022), pp. 124–134.

[27] Q. Zhong, L. Ding, J. Liu, B. Du, H. Jin, and D. Tao, *Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis*, IEEE Transactions on Knowledge and Data Engineering, (2023).

[28] Y. Zhong, P. Chalise, and J. He, *Nested cross-validation with ensemble feature selection and classification model for high-dimensional biological data*, Communications in statistics-simulation and computation, 52 (2023), pp. 110–125.

Edited by: Rajni Mohana
Special Issue on: Sentiment Analysis and Affective computing in Multimedia Data on Social Network
Received: Jul 25, 2023
Accepted: Sep 4, 2023