Sentiment Analysis of Hotel Reviews using Deep Learning

Raksha Muddegowdana Koppalu Prakasha
School of Computing
National College of Ireland
Dublin, Ireland
x19193181

Kenechukwu Otito Ajufo
School of Computing
National College of Ireland
Dublin, Ireland
x19190174

Neil Manoj Bamb
School of Computing
National College of Ireland
Dublin, Ireland
x19149476

***Abstract*—The travel and hospitality industry has seen a vast growth in the past few years and has become one of the prime contributors to the economy. With the advent of social media and web access to the travel and hospitality sector, there has been an abundance of data collected in form of opinions, critiques, experiences, and reviews on the internet. These data collected plays an important role in giving valuable insights into a hotel's or an establishment's future growth. In this paper, we carry out Sentiment Analysis, which aims at classifying these data into positive or negative sentiment. Deep Learning methods such as LSTM (Long Short Term Memory) and BERT(Bidirectional Encoder Representations from Transformers) model have been used to carry out the sentiment classification. Additionally, we will compare the performance of these two models and make a promising choice of model for sentiment classification.**

***Index Terms*—Sentiment Analysis, Neural Networks, LSTM, BERT, Deep Learning**

I. Introduction

Natural Language Processing, (NLP), is a Computer Science (CS) field centred on making human and computer interaction easy and efficient. In its simplest form the system gains an understanding of the syntax and meaning of human language, processes it, and then gives the user output. It entails making a computer system perform meaningful tasks with natural human understandable language as input. Sentiment Analysis, Text Summarization, Information Retrieval etc. are integral parts of NLP.

A crucial aspect of information gathering is being aware of what people think. Sentiment Analysis is used to determine this because sentiments are the most essential characteristic in judging human behaviour. It deals with judgments, responses as well as feelings, which are generated from texts. Sentiment analysis involves the use of a computational approach to derive and identify information and polarity from written text, and this is very relevant given that the internet is a rich resource of sentiment information [1].

Sentiment analysis is utilized in various domains; it is useful in predicting election outcome, stock price, business analytics and other data-driven tasks [2]. Hotel businesses can benefit from the use of sentiment analysis to achieve customer loyalty by understanding the perception of their services from their customers' viewpoints.

Most research on sentiment analysis concentrates on the categorization of overall opinion in text into two classes: positive or negative. This research's major contribution will be the use of deep learning to explore:

- The detection of overall sentiment in the hotel reviews.
- The classification of the hotel reviews into positive and negative classes.

For the purpose of achieving this, a comparative analysis would be conducted on two deep learning models viz Long short-term memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT), to discover how both models fare in accurately predicting the sentiment in the hotel reviews.

The remaining sections in this paper are structured as follows: Section II will detail a general overview of sentiment analysis and then conclude with related works of the applications of machine learning to Sentiment Analysis of user reviews, Section III will discuss components of the methodology, Section IV will present the results and finding of our experiments, and Section V will conclude and briefly discuss the future research considerations.

II. Related Works

The project focuses on the Sentiment Analysis, which is the study and categorization of sentiments in the texts using deep learning techniques [3]. Sentiment Analysis helps businesses to assess their customer's sentiments towards their products and the services. In this section, we will carry out a review of the past works carried out in the field of sentiment analysis.

A. A Survey on Sentiment Analysis

The main purpose of the analysis of sentiments in the context of machine learning and deep learning is to automate a system to recognize and classify the sentiments [4]. With the growth of technology, people have been provided with open platforms to express and share their sentiments and opinions on all of their day-to-day activities. The evolution of social media has also equally contributed to establishing this transparent ecosystem where people from across the globe can share their views and opinions. Rigorous surveys have been carried out in the area of sentimental analysis to portray different tasks, approaches, and applications of sentiment analysis. The survey carried out by Kumar Ravi in [5], focuses on different machine learning and NLP(Natural Language Processing) techniques used for the analysis of sentiment. It also contains an overview of research done in the sentiment analysis field.

Sentiment Analysis is also known as a rapidly growing research area and at times it gets very challenging to keep track of the updates in the said area. To help with the same, a computer-assisted literature review has been proposed in [6] which makes use of text mining and qualitative coding to analyze more than five thousand papers on sentiment analysis from various research journals and similar sources.

Sentiment Analysis is spread over multiple business fields like Marketing, Social Media, Consumer Information, etc and has become a prime aspect of decision making [7], wherein millions of users count on sentiment analysis of opinions shared by others. As per the study, 90% of the customer decisions depend on online reviews, and hence it is important to analyze the sentiments and examine the scores for it [8]. On the contrary, there are also multiple challenges and limitations in the analysis of sentiments at different levels, a paper presented by Ayesha Rashid [9] talks of these limitations and various methods used for sentiment analysis. Finally, the authors in [10] propose approach to extract the opinions from online platforms and discusses the limitations in carrying out Sentiment Analysis on them.

B. Approaches to Sentiment Classification

Sentiment can be an opinion, thought or feelings expressed by an individual, and any kind of analysis carried over the said sentiment is regarded as sentiment analysis. Sentiment Analysis is gaining popularity in the recent trends as the introduction of social media and micro-blogging sites have increased the sentiments/opinions shared by people on these platforms. Sentiments are usually classified into three different classes, positive, negative and neutral.

Tweepy [11] is a proposed solution that extracts sentiment data from twitter and classification algorithms are applied onto the collected data. The feature extraction is carried out using the N-gram techniques and sentiment are categorized as negative, positive, and neutral using supervised classification algorithms of machine learning, k-nearest Neighbour with an accuracy value, greater than 80%. In a similar fashion, [12] discusses sentiment analysis on twitter data, a combination of Sentiword and Naive Bayes were used and it was found out that the combination helped improve the accuracy of classification, by providing negativity, objectivity and positivity score to words present in the tweets. The combination was implemented using Python's Natural Language Toolkit (NLTK) and Twitter Application Programming Interface (API).

Machine Learning and NLP(Natural Language Processing) are the most commonly used approach for sentiment classification. A comparison analysis carried out on these two approaches revealed, Machine Learning techniques evaluated on the parameters of recall, precision and F-measure, provides a high rate of success and accuracy in sentiment classification compared to NLP, but is limited by the domain of data used. A proposal [13] is presented to use ontology in sentiment analysis to achieve a higher rate of success and accuracy. Supervised learning, Lexicon-based techniques, unsupervised learning and hybrid techniques are the most commonly used machine learning based methods for sentiment analysis [14], yet all these techniques are still struggling to deal with complex classification needs.

The limitation of sentiment analysis with respect to NLP is the lack of labeled data, and as a solution to this sentiment analysis is combined with deep learning techniques. The automatic learning capabilities of deep learning make them more favorable to sentiment analysis. [15] describes experiments done to implement different deep learning techniques such as convolutional neural networks, deep neural networks,etc to solve various problems of sentiment classifications like cross-lingual differences, visual-textual analysis, reviews of product, etc.

Despite the extensive use of traditional machine learning classification techniques such as SVM(Support Vector Machines), Naive Bayes, etc. for sentiment classification, a study in recent times promises higher accuracy and success with deep learning techniques. Deep Learning methods of LSTM(Long Short Term Memory) and DCNN(Dynamic Convolutional Neural Networks) applied to Thai tweets [16], returned a result of accuracy that outperformed the traditional machine learning techniques. Recurrent Neural Network [17] applied to Language Modelling also returned promising results as opposed to traditional classification techniques. The Long Short Term Memory (LSTM), part of Recursive Neural Network, helps solve complex, time-lag tasks more efficiently than any other recurrent neural network algorithms [18].

LSTM (Long Short-Term Memory) is a class of RNN designed to deal with temporal sequences and their long term dependencies. Different types of LSTM architecture [19] for sentiment analysis are studied in the analysis of movie reviews. The study started with a simple LSTM model and subsequently, LSTM layer is stacked one upon the other to increase accuracy and finally a Bi-directional LSTM layer was constructed to facilitate data in both forward and backward direction. And it was found that Bi-directional LSTM and stacked LSTM had better performance with an accuracy greater than 83% than the conventional LSTM model.

Based on the bi-directional LSTM model, an attention based mechanism on semantics in the food industry is proposed to recognize feature extraction related to distance. Also, a multi neural network (CNN and BLSTM) for sentiment classification and analysis is proposed in [20] and is characterized by high accuracy and F value.

Bidirectional LSTM(Long Short Term Memory) is used in combination with adaptive weight assignment methods like Symptoms-Frequency Position Attention (BLSTM-SFPA) to carry out sentiment analysis in field of medicine [21]. A hybrid LSTM-CNN Model is proposed in [22] to overcome the problems and challenges of Sentiment Analysis of Movie Reviews. An approach of Word-to-Vector is used for word embedding training. And the results of the hybrid model applied to Amazon movie reviews and IMDB movie reviews and evaluated for the parameters of f-measure, precision, recall and accuracy, outperformed the traditional deep learning techniques. Sentiment Analysis was carried out with 2-pole classification on hotel reviews using RNN(Recurrent Neural Network) [23], and the comparative analysis of the results with traditional classification techniques showed the deep learning method to be more successful.

A combination of CNN and RNN (LSTM) is presented in [24]. The model is built with a bit of hyperparameter tuning, static vectors and then it is applied on a large movie review dataset which resulted in eminent results at multiple sentiment analysis benchmarks.

Another promising model of sentiment analysis using deep learning technique is BERT [25] and is used to solve the fine-grained tasks of sentiment analysis and the results of the model outperform other models used for these tasks even in the absence of refined architecture. BERT Model is also used to carry out Aspect-based sentiment analysis [26] to identify opinion differences towards a product.

A Target Dependent BERT (TD-BERT) model is used to carry out aspect based sentiment classification in [27]. The proposal implements a three target dependent contrasts of base BERT model. The experiment revealed that the TD-BERT model outperforms in terms of accuracy, precision and recall, in comparison to earlier BERT models and embedding based models.

Bidirectional encoders from BERT are used to carry out sentiment analysis of news article to provide insights on the stock market [28]. The model is made to undergo various self learning tasks by pre-training them on a huge volume of domain related documents and articles. The model achieved a F-score of 72.5%.

Aspect-Based Sentiment Extraction and Classification is carried out by applying adversarial training to post-trained BERT model [29]. The proposed model which results in a higher accuracy and precision, outperforms the post-trained BERT model in extracting and classifying aspect based sentiments.

In conclusion, the rationale behind using the LSTM and BERT model is due to the prevalence and performance in the literature. RNN has a problem of vanishing and exploding gradients that fall to zero when transformations are applied. This made RNN very difficult to work with and only worked on short sequences of text. LSTM solves this problem by making long sequence length work; still nowhere near the performance of transformers on a larger text corpus. Transfer learning is not suitable for LSTM, it also is not very reliable in terms of its performance. It was never at par with Imagenet and CNN for transfer leaning. Therefore when LSTM is used, it has to be trained on a task specific labelled dataset from scratch. This changed with the introduction of BERT in 2017.

BERT is selected for this task as the literature suggests it does a better job at understanding the meaning of the context than other NLP and neural networks. Furthermore we conduct a comparative analysis on the performance of the two models.

III. Data Mining Methodology

This section introduces the methods used to build the models. Knowledge Discovery in Databases (KDD) is used for analyzing the dataset in this project. The KDD approach is employed because it makes use of a well structured process. Ultimately, it allows iterative and responsive adjustments to be made in any stage of the process flow.

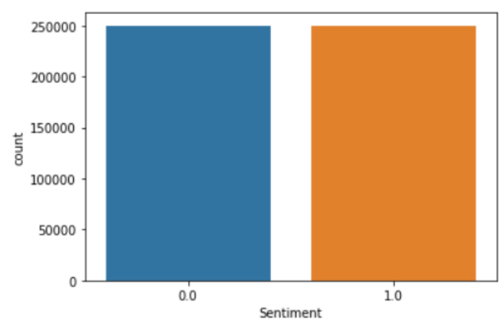
A. Data Selection and Description

The dataset used for the experiment can be obtained from Kaggle [30]. It contains hotel user reviews from the year November 2007 through February 2020. It has 34 attributes and about five hundred thousand instances. Last but not least, the dataset is balanced with equal amounts of negative and positive reviews and the ground truth for these reviews are explicitly mentioned in the dataset.

The original dataset consists of around 500k rows of positive and negative reviews equally. It consists of ”positive-review” and ”negative-review” columns which has positive and negative review of each hotel respectively. Since it was taking too much computation power and running out of RAM in COLAB, we decided to take 250k rows of positive and negative reviews.

B. Exploratory Data Analysis

Exploratory Data Analysis is done to investigate the relations between attributes and instances in the dataset which can not be seen by glancing at the data manually. The simple barchart in Fig 1 shows the numbers of positive and negative reviews in the dataset, while the Fig 2 is a word cloud which is a pictorial image of the most occurring words in the reviews.



(a) Positive and Negative Sentiment Bar chart

(b) Review Word Cloud

Fig. 1: Exploratory Data Analysis

C. Data Preprocessing and Transformation

1) LSTM Model:

In order to obtain better results from the deep learning model, the dataset has to be preprocessed. The positive and negative labels in the dataset are encoded as one and zero respectively, next, the dataset is split in two, 80 percent for training and 20 percent for testing. The reviews are broken down into tokens that are truncated to a maximum length of 200.

The tokenized words are processed using word embedding. Word embedding captures the context of words in the tokenized works in respect to other words in the same space. This process results in the vector representation of each of the tokenized words. The word embeddings are generated with GloVe vectors, which is an improvement over the traditional bag-of-words model encoding schemes. GloVe vectors is a word embedding that is available for free and it contains a hundred dimensional version of the embedding. The subsequent result is the creation of a matrix of one embedding for each word in the dataset.

2) BERT Model:

The data is relatively clean as there were almost zero null or missing values in the data. The dataset had positive and negative reviews for each hotel in separate columns. The preprocessing done at this stage is combining the positive and negative reviews into one single column. The columns are first checked for null values and then appended together into the single column. It is appended in such a way that the first 250k consists of negative reviews and the next 250k consists of positive reviews. Another column was then created which gave it the respective sentiment (ground truth). The ground truth of the data is not manipulated in any way, it is just transformed in a way to make analysis simpler.

D. Data Mining

1) LSTM Model Description:

The LSTM model has seven layers. The first layer consists of an embedding layer which converts the word tokens (integers) into embeddings. The second layer is an LSTM layer followed by a dropout layer. The model also has two dense layers with ReLU activation functions, then another dropout layer, finally the last layer is a fully-connected output layer. The last layer does the job of mapping the LSTM layer output to a desired output. A sigmoid activation function is then used to turn all outputs into a value of either a zero or one. The dropout layers prevent the model from overfitting.

2) BERT Model Description:

BERT is a Bidirectional Transformer that has been successfully tuned and developed for a wide range of Natural Language Processing (NLP) tasks. A BERT model's key task is to produce embedding of terms and sentences (inbuilt pooling) for classification. The main layers are the BERT model followed by the single linear classification layer applied on top of it. The various components / layers are used in the following manner:

- **Input**: The input consists of the hotel reviews in our dataset with respect to the corresponding sentiment (Positive:1 and Negative:0).
- **Tokenization**: The BERT tokenizer converters sentences in the reviews to BERT unique tokens encoding the data. The tokenizer is used to split the raw text into tokens. A token is simply a numerical number that represents a certain word. The training dataset is encoded using a batch_encode_plus method which returns attention masks of each review ensuring the each review are the same dimensionality.
- **BERT-Base Model**: The setting up of the pre-trained model was done using BERT-BASE-UNCASED. It has around 12-layer, 768-hidden, 12-heads and 110M parameters. These are used in combination to generate the final embedding of the output layer. For any BERT model, there is an internal limit of a maximum of 512-word tokens. Since the actual number of tokens for analysis is around 250 tokens, the average token is about 100 tokens. This means that only the first 100-word tokens will be used as the length of the sequence to include more context and details of the reviews.
- **Classification Layer**: The final layer of the BERT model consists of a single linear neural network layer. It is trained from the input layer to the output of the previous layer based on the previous given embedding. It then classifies the input according to the specified classes.
- **Output**: The output generated by the classification layer is either positive or negative sentiment. Here, 1 represents a positive review and 0 represents a negative review.

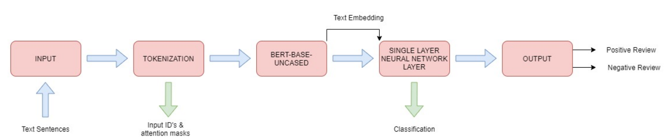


Fig. 2: BERT Model for Sentiment analysis

In order to train the BERT model, it has to be fed in the following inputs: i) input_ids ii) attention_masks iii) labels.

The fine-tuning step is done using the Bert for sequence classification. BERT essentially takes in the text and encodes it in the numerical format based on its own training corpus. A layer of size two is added because the task is to predict two different classes. Next, dataloaders which iterate through our dataset in batches are used. It contains a random sampler in it which prevents the model from overfitting.

The proposed model uses BERT models whose language model has been fine tuned on different domain corpora. The BERT language model is then fine tuned on the corpus containing an equal, balanced classes of hotel reviews. For the BERT language model fine tuning is accomplished with 32 bit floating-point computations using the Adam optimizer. The batch size is set to 32 while the learning rate is set to

$1e-5$. The epsilon is set to the default value of $1e-8$. The maximum input sequence length is set to 256 tokens. A scheduler is also used to schedule high priority tasks as soon as they are ready to be processed. This makes sure all the tasks are run irrespective of their deadlines. It also uses the Adam optimizer and the number of warm up steps as the parameters. The Adam optimizer is set as to the earlier mentioned values and the warm up steps is set to the default value of 0.

TABLE I: BERT model Hyperparameters

Parameter	Value
Dropout Rate	0.1
Batch size	32
Learning rate	$1e-5$
Max epoch	10
Max sequence length	256
Optimizer	Adam

E. Experiment Settings

In this section, the system configuration used for the experiment are described in the table below.

TABLE II: Experimental Environment Details

OS	Windows 10
Memory	16GB
CPU	Intel Core i5
GPU	Nvidia Tesla P100
Software	Google Colab
CUDA version	10.1
Python version	3.6
Keras version	2.4.3
TensorFlow version	2.3

IV. Evaluation

The accuracy, precision, recall and F1 Score are used as the performance metrics to evaluate the models.

- **Accuracy**
  Accuracy is used to measure the proportion of the total sentiment predicted that were correct. It is an important evaluation metric for the model as there is no class imbalance in the dataset. It is calculated using formula 1,

$$Accuracy = \frac{NumberofCorrectSentimentPredictions}{TotalNumberofPredictions} \tag{1}$$

- **Precision**
  Precision is used to measure how precise a model is in predicting the right sentiment and answers the question, on the percentage of our results that are relevant. It is calculated using the formula 2,

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{2}$$

- **Recall**
  Recall is used to measure the percentage of actual correct right sentiments the model identified(True Positive) in the presence of falsely identified sentiment.It is calculated using the formula 3,

$$Recall = \frac{FalsePositive}{TruePositive + FalseNegative} \tag{3}$$

- **F1 Score**
  F1 Score is an average of Precision and Recall and takes into consideration both False Positives and False Negatives. F1 plays an important role in evaluation of models in the case of unevenly distributed classes. F1 Score is calculated using formula 4,

$$F1Score = \frac{2 * (Recall * Precision}{Recall + Precision} \tag{4}$$

A confusion matrix computed in Figure 3 gives the values of True Positives, False Positives, True Negatives, and False Negatives produced by a model which is used to calculate the above matrices.

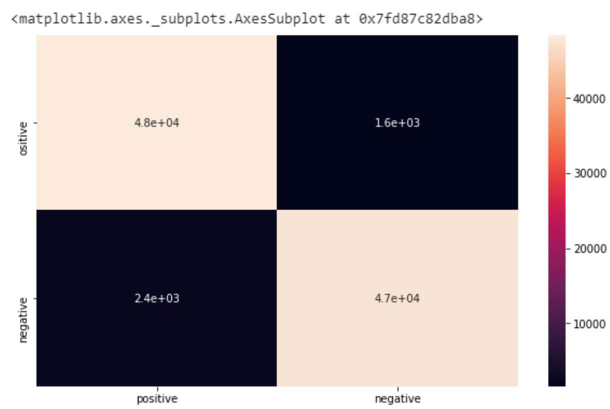


Fig. 3: Confusion Matrix for LSTM Model

The table III gives an overview of performance metric values calculated for sentiment classification using LSTM and BERT models.

TABLE III: Model Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
LSTM	95.96	96.99	94.82	95.89
BERT	97.65	97.59	95.72	96.68

The Figure 4 gives the overview of the classification Report.

```
The Classification report is::
              precision    recall  f1-score   support

           0       0.95      0.97      0.96     50152
           1       0.97      0.95      0.96     49696

    accuracy                           0.96     99848
   macro avg       0.96      0.96      0.96     99848
weighted avg       0.96      0.96      0.96     99848
```

Fig. 4: Classification Report for LSTM

The BERT model gives a better accuracy to the LSTM model with fewer parameters and embeddings. For evaluation, a function is created which trains the model for the training set and tests it on the validation set. The validation set is the same as the training set except it does not perform back-propagation. The evaluation mode essentially freezes all our weights and there are no gradients associated with it. This gives us an accuracy of about 97%.

V. Conclusion And Future Work

Reviews serve a crucial role in modern day businesses as they can be the determinant factor for a customer deciding to buy a product or service from the business. The gains in information technology afford human beings to express their feelings, thoughts and emotions easily, as a result, reviews are all over the internet. Businesses therefore exploit the information to gauge how the service or good they provide are perceived by the general public.

In this research, we attempt to do a sentiment analysis on hotel reviews using two deep learning techniques - LSTM and BERT. The objective here is to discover how well these models perform in predicting the overall sentiment of users' reviews. To evaluate the models' performance, the accuracy, precision, recall and F1 Score are used. From the experiments conducted the LSTM model had an accuracy, precision, recall and F1 Score of 96%, 97%, 95% and 95% while the BERT model had 97%, 98%, 96% and 97% respectively.

In conclusion, future research can explore aspect-based, emotion and spam and fake detection sentiment analysis. These can be implemented based on appropriate datasets.

References

[1] X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 1, p. 5, 2015.

[2] D. Li, R. Rzepka, M. Ptaszynski, and K. Araki, "Hemos: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media," *Information Processing & Management*, vol. 57, no. 6, p. 102290, 2020.

[3] S. Padmaja and S. S. Fatima, "Opinion mining and sentiment analysis-an assessment of peoples' belief: A survey," *International Journal of Ad Hoc, Sensor & Ubiquitous Computing*, vol. 4, no. 1, p. 21, 2013.

[4] D. Osimo and F. Mureddu, "Research challenge on opinion mining and sentiment analysis," *Universite de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment*, vol. 508, 2012.

[5] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14–46, 2015.

[6] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—a review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16–32, 2018.

[7] T. Chalothom and J. Ellman, "Simple approaches of sentiment analysis via ensemble learning," in *information science and applications*. Springer, 2015, pp. 631–639.

[8] L. Peng, G. Cui, M. Zhuang, and C. Li, "What do seller manipulations of online product reviews mean to consumers?" 2014.

[9] A. Rashid, N. Anwer, M. Iqbal, and M. Sher, "A survey paper: areas, techniques and challenges of opinion mining," *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 6, p. 18, 2013.

[10] B. Seerat and F. Azam, "Opinion mining: Issues and challenges (a survey)," *International Journal of Computer Applications*, vol. 49, no. 9, 2012.

[11] P. Tyagi and R. Tripathi, "A review towards the sentiment analysis techniques for the analysis of twitter data," in *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*, 2019.

[12] A. Goel, J. Gautam, and S. Kumar, "Real time sentiment analysis of tweets using naive bayes," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*. IEEE, 2016, pp. 257–261.

[13] M. R. Yaakub, M. I. A. Latiffi, and L. S. Zaabar, "A review on sentiment analysis techniques and applications," in *IOP Conference Series: Materials Science and Engineering*, vol. 551, no. 1. IOP Publishing, 2019, p. 012070.

[14] Z. Madhoushi, A. R. Hamdan, and S. Zainudin, "Sentiment analysis techniques in recent works," in *2015 Science and Information Conference (SAI)*. IEEE, 2015, pp. 288–291.

[15] Q. T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat, and A. Rehman, "Sentiment analysis using deep learning techniques: a review," *Int J Adv Comput Sci Appl*, vol. 8, no. 6, p. 424, 2017.

[16] P. Vateekul and T. Koomsubha, "A study of sentiment analysis using deep learning techniques on thai twitter data," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2016, pp. 1–6.

[17] T. Mikolov, M. Karafiát, and L. Burget, "Jan ˇcernocky, and sanjeev khudanpur. 2010. recurrent neural network based language model," in *Eleventh annual conference of the international speech communication association*, 2010, pp. 1045–1048.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[19] S. Pal, S. Ghosh, and A. Nag, "Sentiment analysis in the light of lstm recurrent neural networks," *International Journal of Synthetic Emotions (IJSE)*, vol. 9, no. 1, pp. 33–39, 2018.

[20] Y. Li, Q. Jin, M. Zuo, H. Li, X. Yang, Q. Zhang, and X. Liu, "Multi-neural network-based sentiment analysis of food reviews based on character and word embeddings," *The International Journal of Electrical Engineering & Education*, p. 0020720920928492, 2020.

[21] M. Bi, Q. Zhang, M. Zuo, Z. Xu, and Q. Jin, "Bi-directional lstm model with symptoms-frequency position attention for question answering system in medical domain," *Neural Processing Letters*, vol. 51, no. 2, pp. 1185–1199, 2020.

[22] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, "A hybrid cnn-lstm model for improving accuracy of movie reviews sentiment analysis," *Multimedia Tools and Applications*, vol. 78, no. 18, pp. 26 597–26 613, 2019.

[23] K. KARAOĞLAN, V. TEMİZKAN, and O. FINDIK, "Sentiment analysis for hotel reviews with recurrent neural network architecture."

[24] A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *Ieee Access*, vol. 6, pp. 13 949–13 957, 2018.

[25] M. Munikar, S. Shakya, and A. Shrestha, "Fine-grained sentiment classification using bert," in *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, vol. 1. IEEE, 2019, pp. 1–5.

[26] C. Sun, L. Huang, and X. Qiu, "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence," *arXiv preprint arXiv:1903.09588*, 2019.

[27] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-dependent sentiment classification with bert," *IEEE Access*, vol. 7, pp. 154 290–154 299, 2019.

[28] M. G. Sousa, K. Sakiyama, L. de Souza Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "Bert for stock market sentiment analysis," in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2019, pp. 1597–1601.

[29] A. Karimi, L. Rossi, A. Prati, and K. Full, "Adversarial training for aspect-based sentiment analysis with bert," *arXiv preprint arXiv:2001.11316*, 2020.

[30] A. D, "Trip advisor reviews," Mar 2020. [Online]. Available: https://www.kaggle.com/arthurdfr/trip-advisor-reviews