Improving the Performance of Aspect Based Sentiment Analysis Using Transformer Based Techniques

MSc Research Project
Data Analytics

Prathamesh Gaikar
Student ID: x19216301

School of Computing
National College of Ireland

Supervisor: Dr. Bharathi Chakravarthi

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Prathamesh Gaikar
Student ID:	x19216301
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr. Bharathi Chakravarthi
Submission Due Date:	16/08/2021
Project Title:	Improving the Performance of Aspect Based Sentiment Analysis Using Transformer Based Techniques
Word Count:	6930
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	22nd September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Improving the Performance of Aspect Based Sentiment Analysis Using Transformer Based Techniques

Prathamesh Gaikar
x19216301

Abstract

In today's highly competitive market, word of mouth from buyer's viewpoint has been vital step towards success for any company. Currently, most of the enterprise in all the sectors have launched their websites to sell their products and services. Each day, millions of reviews, opinions, and sentiments are generated on the online websites regarding products and services. It is very challenging to handle and comprehend such large amount of opinion based data. Sentiment analysis is the domain which acknowledges and extracts the emotions from available opinioned data and analyze the process through natural language processing and text classification. It is the method to understand the sentiments in the text and is also becoming challenging in many research areas including the data mining field as there is a rapid increase in the number of web pages which includes product and service reviews. In this research SpaCy library is used to extract the aspect terms. For the first part of the research, transformer based models are applied without aspect based method and for the second part aspect based method is applied to same models and their results are evaluated and compared. The experiments result shows that the accuracy of transformer based models without aspect based approach is bit higher than the one with the aspect based method.

Keywords - - Sentiment Analysis, Natural language processing, Aspect based sentiment analysis (ABSA), Transformer based pre-trained models, text classification, XLNet, RoBERTa, ALBERT, SpaCy, Textblob.

Contents

1 Introduction **4**
1.1 Overview . . . 4
1.2 Motivation . . . 5
1.3 Research Question . . . 5
1.4 Research Objectives . . . 5
1.5 Contribution . . . 5

2 Related Work **6**
2.1 Significance of Text Classification . . . 6
2.2 Review of Aspect based Sentiment Analysis . . . 7
2.3 Transformer Based Approach . . . 8
2.4 Pretrained models . . . 8
2.5 Aspect Extraction . . . 10

3 Methodology **11**
3.1 Business Understanding . . . 11
3.2 Data Acquisition . . . 12
3.3 Data Pre-processing . . . 12

4 Design Specification **14**

5 Implementation **14**
5.1 Environment Setup . . . 14
5.2 Data Preparation . . . 15
5.3 XLNET . . . 15
5.4 ALBERT . . . 16
5.5 RoBERTa . . . 16

6 Evaluation **17**
6.1 Experiment with XLNet . . . 17
6.2 Experiment with ALBERT . . . 17
6.3 Experiment with RoBERTa . . . 17
6.4 Experiment with XLNet based on Aspect terms . . . 18
6.5 Experiment with ALBERT based on Aspect terms . . . 18
6.6 Experiment with RoBERTa based on Aspect terms . . . 18

7 Discussion **19**

8 Conclusion and Future Work **20**

List of Figures

1	Example of Aspect Based Sentiment Analysis	4
2	Overview of CRISP-DM model	12
3	Star Ratings distribution	13
4	Labelled Sentiments Plot	13
5	Aspect average Score Plot	14
6	Design Specification	15
7	XLNet Factorization referred from Yang et al. (2019)	16
8	Confusion matrix of RoBERTa	17
9	Confusion matrix of RoBERTa with aspect terms	18
10	Comparison of XLNet with and without aspects	19
11	Comparison of ALBERT with and without aspects	19
12	Comparison of RoBERTa with and without aspects	19

List of Tables

1	Data Structure	15

1 Introduction

1.1 Overview

Natural language processing (NLP) is one of the key techniques in promoting sentiment analysis. It one of the areas of artificial intelligence which empowers the computers to understand varieties of languages spoken by humans. For sentiment analysis, NLP concentrates on recognizing and managing the unstructured data present in e-commerce websites. To bring inferior impression about the products or services, NLP based approach will be used to analyze the sentiments.
The text analysis method in which data is categorized by aspect and attributed sentiment to each one is identified is Aspect based sentiment Analysis. Customers feedback can be analyzed by associating different aspects with specific sentiments. ABSA is different from document-level and sentence-level sentiment analysis and for each aspect of given product or service can gain more precise interpretation of sentiments. ABSA can extract two things i.e., sentiments and aspects. ABSA system receives input as a set of texts discussing a particular entity. The main aspect of the entity is detected and average sentiment of text per aspect is estimated by these systems.Fig 1 shows an example of Aspect based sentiment analysis.

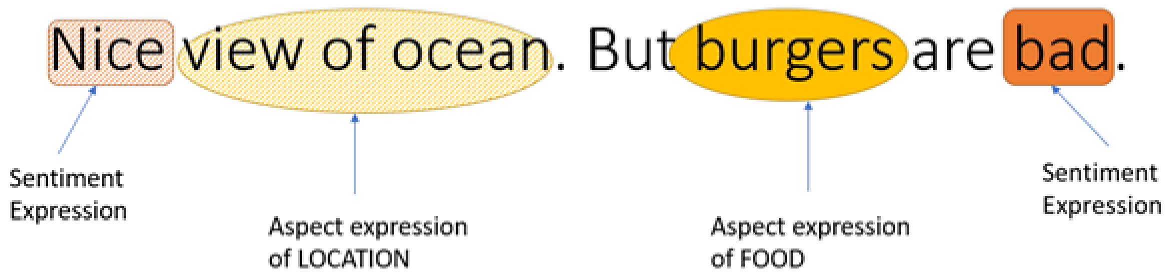

Figure 1: Example of Aspect Based Sentiment Analysis

One of the giant e-commerce platforms used by millions of people all around the world is Amazon. According to the research it is revealed that over 88% of the people trust online reviews. These trends and opinions of consumer can possibly lead to better stock market prediction. Any product with positive reviews provides a powerful explanation of the product. More the number of reviews, it is more encouraging for other online shoppers.
Advancement in the field of NLP has led to growth of pre trained NLP models. Models such as OpenAI's GPT-3 and embeddings from Language model (ELMo), ULMFiT, ALBERT are trained with huge volumes of data, which results in better results when trained with less data. Surpassing Vaswani et al. (2017) other neural network like convolutional neural network and recurrent neural network, transformer based has become a influential architecture for natural language processing. In this research, we will be first applying Aspect based sentiment analysis technique on the dataset and then applying the pre-trained models to it. Secondly, we will be applying pre trained NLP models directly to the dataset and then will be comparing all the models based on performance.

1.2 Motivation

For the past decade, online businesses have increased and gained popularity. With the advent in Web 2.0, enormous amounts of data generated every moment. The dimension of the world is shrunk since social media has taken over. Millions of consumers shop online, and they share their opinions about the purchased product or service. The information provided by the consumer influence others. With large number of customer review on the webpages or applications, information can be retrieved regarding the customers' sentiments. To deal with market competitors, customer's sentiment is one of the vital aspects for renovating enterprise strategies. To obtain knowledge about the consumer's opinions of their product is the need of many organizations. The stage of obtaining the reviews and opinions about the products or services has made the internet a substantial source of getting views about the product based on positive and negative comments. In an opinioned data Shahnawaz and Astya (2017), the process of recognizing and classifying the expressed sentiments, to know the attitude of consumer towards a particular product is known as sentiment analysis. People's opinion, attitude and emotions towards product or services can be studied computationally. Sentiment analysis is used to get useful insights from an unorganized and unstructured data from various online platforms such as e-commerce, social media, etc. Therefore, reviews provide useful views based on the product, service, or company. This information can be used by the businesses to gain the insights and improve the quality of the product and services. Thus, it is very crucial for every E-commerce business to evaluate and analyze such content.

1.3 Research Question

"Can Aspect Based Sentiment Analysis (ABSA) technique with the obtained aspect terms improve the performance of pre-trained models like XLNet, ALBERT and RoBERTa in comparison to the same models without the ABSA technique for Sentiment Classification using Amazon data?"

1.4 Research Objectives

To get the desired results, research objectives that were carried out are given below:

- Identifying the attributes of product that people are commenting on. (Also known as Aspects)
- Using the extracted attributes to classify sentiments of the review.
- In first part XLNet, ALBERT, RoBERTa will be implemented without using Aspect based techniques.
- In second part, we will be implementing XLNet, ALBERT, RoBERTa with Aspect based techniques.

1.5 Contribution

Combining Aspect based sentiment analysis technique with the state of the art pretrained models is the one of the contributions of this research. The research shows that pre trained transformer models such as XLNet, RoBERTa and ALBERT improves upon performance of the model and also achieves good results in Natural Language processing. Another major contributions of the research is to provide a framework to extract the aspects from the review and label the sentiments. This aspect terms are extracted using spaCy library and textblob. On domain dependent data it works well. By extracting aspects from any data this research has created path for researchers to implement aspect based sentiment analysis.

2 Related Work

This section provides a thorough literature review of the terms involved in Natural language processing, sentiment analysis and existing research which has been done in this field. Further, transformer based deep learning methods and aspect based techniques are also considered in detail

2.1 Significance of Text Classification

The number of online texts has exploded over the years with the advancement of internet. So, the retrieval of Knowledge from this text has attracted many researchers and this retrieval is based on text classification. It computes category Faguo et al. (2010) of induction by means of induction from a large scale corpus, so it is an inductive learning. In this paper the focus was to text classification and its further practical application such as analysis of the sentiments or the aspects of that text. In the field of web text information processing, it is widely used and successfully applied.
According to Ma et al. (2018) the development process of text classification is classified in 3 stages: i) Exploration phase, ii) Theoretical research phase, iii) Application practice phase and this research defines techniques and development process of text classification which includes text representation, classification of algorithms and feature selection. These results in advantages and disadvantages of current mainstream classification technique. One of the important tasks of text classification is feature extraction since the textual data is unstructured. In the field of e-commerce and log message analysis, text classification is widely used and an essential one for tasks such as text processing , news classification, sentiment classification Zhang and Rao (2020) and in this research a new approach is proposed to convert the natural text to feature like bag of words with n-gram technique.
In this research extraction of feature words based on chi-square statistics has been proposed by Zhai et al. (2018). Here, the single word and double word technique as feature is used to classify text at the same time. In this paper model for text classification is proposed by Yao et al. (2020) based on fastText and the researcher has anticipated emotional polarity judgement task. For this research results are showing that precision, recall and f1-measure are superior to model based on traditional machine learning methods. This technology of text classification Yao et al. (2020) solves the issues caused by vast text data by conveniently detecting the necessary resources and enhance the efficiency of data utilization.

2.2 Review of Aspect based Sentiment Analysis

The main aim of sentiment analysis is to classify emotion involving text as positive, negative, and neutral. There are three types of level in sentiment analysis namely : Document level, Sentence level and Aspect level. In document level sentiment analysis, the polarity is figured out by analyzing the whole document. In this analysis a file containing reviews is fully analyzed and the polarity report is produced. On a single entity, document expresses the whole opinion. According to Shirsat et al. (2017) it is helpful only when we want to know the overall polarity of a particular entity, but its disadvantage is that we will not know what people exactly liked or disliked about that product also it is not applicable on multi product reviews entity. In this paper, proposed by Shirsat et al. (2017) document level analysis has been done to find the polarity of the news.

In sentence level analysis the aim is to determine the sentiment of each sentence i.e., positive, negative, or neutral sentence. Generally, neutral has no meaning. Subject Zhang and Liu (2017) classification is closely related to this analysis since it differentiates accurate information that is expressed from the sentences from the subjective views and opinions. As many objective sentences refers to opinions, subjectivity is not equivalent to sentiment. In this research paper by Basiri and Kabiri (2018) based on uninorm operators a new sentence level accretion mechanism is proposed. For polarity detection and score prediction the mentioned method was applied on four large datasets to show the utility. To show the superiority of aggregation method, performance of aggregation method was compared to performance of Dempster- Shafer aggregation method ad the results were showing that the proposed method outperforms the DS method.

The aim of the Aspect based sentiment analysis is to first extract the aspects from the reviews or sentences and then determine their polarities of sentiments. In general, different tasks Wang et al. (2021) are involved in sentiment analysis such as objectivity identification, polarity classification and multimodal fusion and in this paper, researchers have focused on determining the polarities of aspects by using sentHint framework for aspect level sentiment analysis which integrates DNN and linguistic hints in coherent MLN inference model.

In Aspect based sentiment analysis the final classification sometimes may be affected by sentence structure, so the researchers Yang and Yang (2020) proposed a method based on gating mechanism with neural network and self-attention mechanism to solve this problem. So, in this paper firstly the self-attention mechanism is used to extract the structural features of the text and then integrate it with the features extracted by original sentence with gated CNN and then combine it with aspect terms to finalize sentiment feature. Here, for the performance comparison CNN, In research by Zhao et al. (2021) a model combination of CNN and grated recurrent unit is proposed for aspect based sentiment analysis. In this local feature created by CNN are utilized and it shows that the proposed methodology achieves tremendous performance in terms of sentiment classification and aspects extraction.

The method should be more unsupervised than supervised to work in various domains so it can become independent of the area of work. In this research Do et al. (2019) precise approach of major deep learning techniques for aspect level sentiment analysis is proposed. Common approaches of CNN, RNN,LSTM,GRU with their variants were implemented to analyze, classify, understand, and predict nature of polarity of the sentences or text to fine grained levels.

2.3 Transformer Based Approach

Attention mechanism introduced by google was presented as a solution to problem faced by LSTM model. The basic concept behind this attention mechanism was to determine which particular words are most related to its paragraph overall meaning. The problem of fixed internal context vector is solved. Internal vector now shares all the hidden layers. In this mechanism, the performance depends on overall input state on which the model has paid the attention and not only last state. In field of neural machine translation and entity recognition transformer based techniques has demonstrated an excellent performance and is also popular in deep neural network model for neural language processing. It requires hardware acceleration resolution and is computationally and memory intensive. During inference, its scaled attention mechanism in auto regressive decoder brings a performance blockage. In this paper by Yang et al. (2020) ReTransformer technique is proposed which not only eliminate the data dependency by avoiding the writing result using proposed matrix decomposition by using mat- mul operations but also try to accelerate the attention mechanism. Transformer is comprised of similar model structure shared by decoder stack and encoder stack.

The field of Natural Language Processing has got huge attraction from the researchers with the introduction of transformer based language models. In this research Pipalia et al. (2020) Imdb reviews dataset is used for showing the classification strength of emotions from pre trained models and what amount of accuracy can be achieved by using these models. In Natural language processing, transformer is architecture that holds the extended span control and focusses to solve sequence to sequence problems. The decoder and encoder have same structure except one significant change, in decoder to manage key and value vectors forwarded from encoder layer at each phase there is a special Encoder-Decoder attention layer. Also, a query vector is retained by decoder which was generated by self-attention layer. The reason of training data is to predict the next term, so the subsequent words are masked. The input is not treated in series by the decoder so words in structure are encoded and while translating the words into embeddings, positional data is added which facilitates transformer model to consider sequence order of words.

2.4 Pretrained models

In the field of computer vision, pre-trained techniques were firstly proposed. Recently they have been used for various tasks in the field of natural language processing. To improve the classification ability of sequence to sequence model Google team and value of demand of pre-training are put into view again. In this paper Duan et al. (2020) have given introduction to development and achievement of pre-trained language model. Several pre-trained language models were discussed such as BERT, variations of BERT, etc and will help other learners to explore and understand pre-trained models for various natural language processing and also the advantages and disadvantages of each model were analyzed.

i) XLNet

For the wellness of society and individual wellbeing, modelling pessimism and optimism precisely in social media is an important function. In this paper Alshahrani et al. (2020) multiple models are built on top of XLNet for prediction of cynicism and optimism on twitter messages and positive emotions are much more common in optimistic messages while in pessimistic messages negative emotions are more is demonstrated using XLNet sentiment analysis. Here, XLNet model is fined tuned to predict both user level and in-

dividual level outlook emotions in twitter messages. Using multi-head attentions XLNet model can secure right and left perspective jointly and calculate contextualized symbols. Twitter messages data include special characters which must be handled correctly since it has some existing meaning even then the model gave the top accuracy on this dataset. Applying knowledge obtained from the source field to train the target field is the aim of cross-domain sentiment classification. In this research by Myagmar et al. (2019) has proposed BERT and XLNet models for classification of cross domain sentiments and discover the transferability in perspective of cross domain through in complexity analysis of two models performance and update novelty with considerable scope of enhancements. For cross domain sentiment classification task both BERT and XLNet outperforms the previous state-of-art. With using around 50 fined tuned samples XLNet achieves the state-of-art results by efficiently capturing the context. Compared to BERT, XLNet is more resource voracious and also learns contextual data with a few numbers of fine tunings.
Broad research has been done on sentiment analysis for software engineering (SA4SE). In this research by Zhang et al. (2020) has proposed and evaluation of existing SA4SE tools and pre-trained models in which the research was as compared to SA4SE tools, how accurate and efficient is pre-trained transformer models. Here, XLNet has predicted 78% of the reviews correctly and XLNet generates better results for long text by integrating relative encoding scheme and segment recurrence mechanism of transformer and outperforms BERT in sentiment analysis. Firstly, the effectiveness of pre-trained model is monitored on the sentiment analysis for software engineering task is examined in which XLNet outperforms SA4SE tasks.

ii) ALBERT

Downstream task is enhanced when model size is increased for training natural language but due to limit of GPU memory and longer training time, we cannot increase the model size after a certain point. To solve this issue in this paper Lan et al. (2020) , parameter reduction technique is proposed, and this proposed method shows that it is more scalable and accurate than the original BERT i.e., in this research for the all the discussed problems A Lite BERT (ALBERT) architecture is designed which has less parameters than the original BERT. The major hurdle in scaling pre-trained model is boosted by integrating two parameter reduction technique. The two techniques are factorized embedding parameterization and cross-layer parameter sharing. By applying both the methods, number of parameters are reduced without affecting the performance of model. In total, ALBERT gets much better results than BERT. Traditional word vector model is used by existing sentiment analysis product review model based on deep learning in which it is difficult to obtain contextual semantic information of words. In this research by Wang et al. (2020) a model of ALBERT and LSTM is proposed for sentiment analysis of product review. Here, first the pre-trained ALBERT model is used to get the word vector which contains positional and semantic data and LSTM model is used to obtain semantic features for training. So, this syntactical structure and semantic data makes the accuracy of model much better than the other word vectors which develops product review sentiment analysis.
In paper by Zhang et al. (2020) has proposed evaluation of pre trained models and sentiment analysis for software engineering. In this evaluation is conducted based on considering five sentiment analysis for software engineering tools and pretrained models and along with this there are six different datasets. For this research along XLNet, ALBERT

is also used for evaluation and all pre-trained models outperformed the SA4SE. For this ALBERT base V1 model is used with 11M parameters, 12 layers, 768 hidden layers and 12 multiheads. On API and SO datasets ALBERT is the best pre-trained model with micro and macro f1 score of 82% and 89% respectively.

iii) RoBERTa

In the current pandemic situation, the significance of newspapers has become more critical and very important. In this research Ghasiya and Okamura (2021) the used database has more than a lakh headlines and it was analyzed using word2vec and RoBERTa model. Here, RoBERTa model has accomplished a accuracy of 90% and the headlines were classified better than the traditional classifiers and this execution of RoBERTa for sentiment classification on the dataset showed that the 73.23% of UK news had negative sentiments, while South Kore news has 54.47% of positive sentiment.The reason behind the better performance of RoBERTa method is that it considers context for each incidence of provided word. It is trained on much longer sequences and larger datasets with dynamic mask generation and without the objective of next sentence prediction.

The impact of hyperparameters and training data size must be monitored carefully which is proposed by Liu et al. (2019) in this paper. In this replication study of BERT is presented in which we find that BERT model is undertrained and exceeds the performance of every model and the results are highlighted. Here, we observe that RoBERTa offers large enhancements in results over BERT. When huge number of data is trained over RoBERTa, it shows further enhancements in performance across all downstream tasks.

In paper by Pipalia et al. (2020) researchers have proposed method to examine the classification strength from pre-trained models. Along with different pre-trained models, RoBERTa is also used in this research where it has more than 50 thousand sub-words and the masking method is dynamic. RoBERTa model is applied with a base size of 110 and large size of 340 and BERT without next sentence prediction is used which gives experimental accuracy of 94.2%.

Transformer based methods have a better understanding in natural language processing. In this paper Adoma et al. (2020) efficiency of various pre-trained models is compared in identifying emotions from text. Here, In this experiment RoBERTa has highest accuracy of identifying emotions on ISEAR dataset over the other pre-trained models considering their precision, recall and F1-measure.

One of the fastest growing field of artificial intelligence is Natural language processing. In this research by Pham et al. (2020) a pipeline to adjust RoBERTa language model to a Vietnamese hate speech detection is proposed. Here, RoBERTa model has been demonstrated by intensively performing experiments using PhoBERT, where fine tuning strategy of RoBERTa is extremely efficient in task of text classification. In this experiment various training methods are used such as block wise learning rate, layer freezing and label smoothing. Also, it indicates that the proposed method of pipeline boosts the performance substantially and achieving new state of art by Vietnamese hate speech detection with an F1-score of 0.7221.

2.5 Aspect Extraction

Aspect extraction is identification of key terms referred to important features of the targeted product or service along with the associated positive and negative emotions. Typically, there are several aspects and sentiments in the same sentence. Significant traction is gained by spaCy library in performing sentiment analysis. In this research Soni and Rambola (2021) tasks such as aspect detection from the emergence of deep learning has motivated to use implicit aspect detection. Also, it is noticed that efficiency is significantly improved by using WordNet. Hence, a hybrid model of RNN trained on dataset and similarity calculations based on wordnet and similarity function based on spaCy is proposed to detect the implicit aspects. When people understood that the whole text may have different sentiments in relation to different entities due to which need of aspect based sentiment analysis was raised. In this research Singh and Verma (2021) the model is trained on reviews and then the sentiments and aspects are mines using spaCy, textblob and Vader library. The whole intention and main focus of the research was to speedup the process of aspect based sentiment analysis and the aspects mining process should be easy. Here the model has worked exceptionally well, and it is efficient and effective. In this the results the result with the aspect sentiments were more accurate than the text sentiments. For getting more better results, this approach can be used on large datasets. After extracting the opinions and sentiments better results can be achieved with both labelled and unlabeled data. The technique of easy extraction of required news events is called as topic identification. In this research C.P. and Sunitha (2020) to make information extraction efficiently work for news events, a method is proposed in which spaCy library is used to identify 16 NER tags. spaCy library used conditional random field and Hidden Markov Model to identify the tags. Also, part of speech and Named Entity Recognition is used. Accuracy of the 16 NER tags is calculated using different supervised models by using spaCy.

3 Methodology

In this section the methodology used in this research is discussed. The most popular used methodology is CRISP-DM and we have used it in our research. Here the methodology is represented in 4 main stages i.e., from understanding of data, preprocessing, modelling and evaluation of transformer based models. For effective decision making, the evaluated results could be used.

3.1 Business Understanding

To sort the customer conversations, tweets or reviews manually is the big challenging task. Also analyzing the data at the granular level is not possible. And also, humans are not always objective due to some personal experiences and beliefs and always agree around 60-65 of times for concluding sentiments from the text. For analyzing the data automatically aspect based sentiment analysis is used by the businesses since it saves time and money. To gain the insights and improve the quality of product or service, reviews can be used. And by applying a centralized model the same conditions can be applied to all the texts so that the results will accurate and consistent. To evaluate and analyze such reviews it is very crucial for every business. The main problem is to know the underlying problem. Rather than spending time in analyzing each sentence, its easier to scan and categorize text as positive and negative. Also, a deeper understanding of specific product and services can be gained easily due to which the main focus will be on customers' requirements.

Figure 2: Overview of CRISP-DM model

3.2 Data Acquisition

The data is gathered from official website of Amazon web service. The data was downloaded in .tsv format. The dataset consists of 1705837 records. The data has various parameters such as customer id, review id, product id , review body, review title, star ratings, product parent, product title, product category, helpful votes, total votes, vine, review headline, review date. The dataset has reviews of different products from the year 2004 - 2015. To achieve the objective and goal of the project, the data has to be polished, and models should be trained on data.

3.3 Data Pre-processing

Crucial steps such as checking the missing values, detecting the languages, labelling the data based on star ratings and detecting the aspect terms and labelling them using spaCy library were done in this step. The dataset consists of 1705837 records. Since the data is very huge, for preprocessing first 1000000 data is used. There were no missing values in the dataset. For detecting the language, Langdetect package was installed. Since in this only English language model of spaCy library is used, Langdetect is used for filtering the English reviews. After filtering the English reviews, the reviews in other than English language were not used in the research. Data labelling is one of the important steps of preprocessing for machine learning. If the data is labelled properly than it provides ground truth that the machine learning algorithm uses to check its prediction accuracy. In this data has been labelled on the basis of star ratings. star ratings have values from 1 to 5 of which the distribution is shown in figure 3. So, whenever the star ratings value is above 3 then it is considered as positive sentiment and whenever the star rating is lower or equal to 3 then it is considered as negative sentiment as shown in figure 4. The positive sentiment is denoted as '1' and the negative sentiment is denoted as '0'.
Another important step in this research in preprocessing is aspect extraction. For aspect

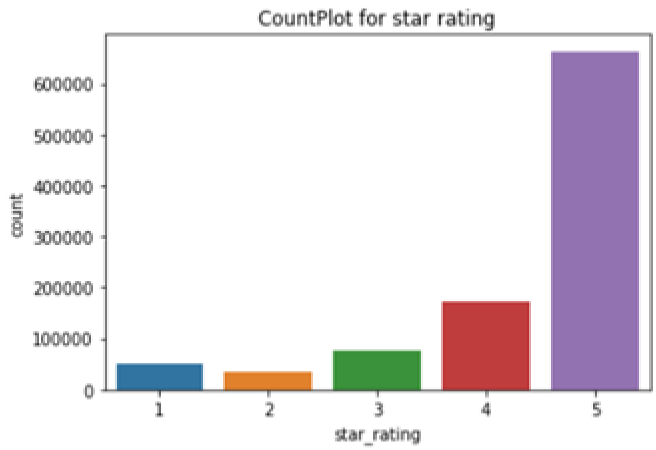

Figure 3: Star Ratings distribution

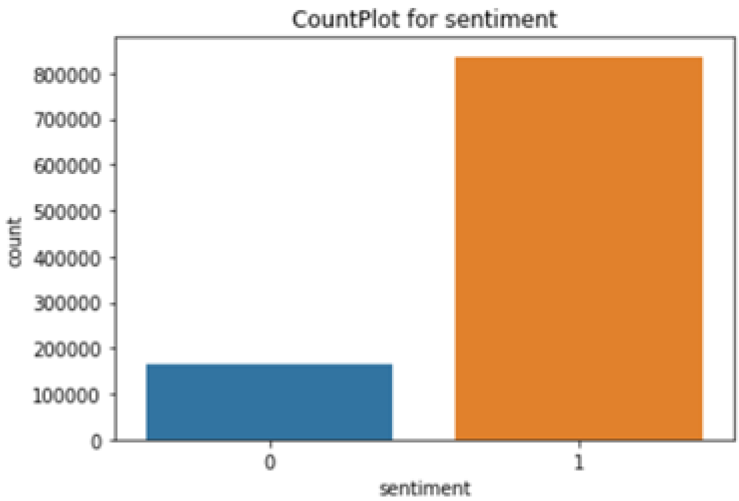

Figure 4: Labelled Sentiments Plot

extraction, spaCy library is used. Here after extracting the aspect terms and calculating the sentiments as positive or negative, a column has been added to the data of aspectAverageScore. In this data, for calculating aspectAverageScore first the aspect terms and the sentiments associated with it are extracted. If there are more than 1 aspect and associated sentiment, then the average of the sentiment is taken. For example. The book title is good, but the stories are bad and also the pictures are not good; In this example book, stories and pictures are the aspects and there are associated sentiments with it. So according to the average of the sentiment it is marked as negative sentiment i.e., 0 in aspectAverageScore column in the dataset and graphically is represented in figure 5

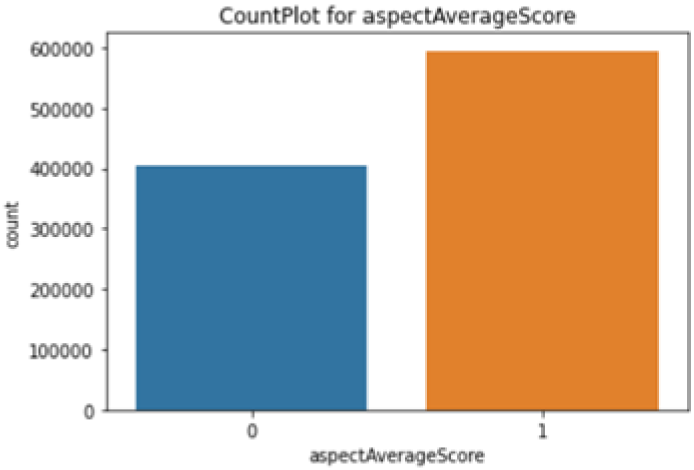

Figure 5: Aspect average Score Plot

4 Design Specification

The architecture followed for this research were as follows:

- At first, the dataset was downloaded from amazon web services website.
- Then the dataset was loaded in google colaboratory, where it was initially preprocessed. Some histograms were plotted for getting valuables insights. Also, the unnecessary data was removed. The reviews which were in other languages than English, were removed from the dataset since the spacy library is trained on 33333different corpus. Depending on the star rating the review is marked as positive or negative. spaCy library is used to detect the aspect terms from the review body.
- Data is splitted into train and test in 70:30 ratio respectively followed by implementation of transformer based models i.e., ALBERT, XLNet and RoBERTa.
- These models are evaluated on the basis of precision, recall, f1-score, and classification accuracy.
- Finally, the results were visualized in the form of histogram using python libraries.

5 Implementation

The implementation consists of following steps: Environmental setup, Data Preparation and implementation of models.

5.1 Environment Setup

Implementation of this research was performed on Google Collaboratory. It is online platform build upon Jupyter notebook and its is free. It allocates to run python programs on google servers and influences high end GPU's free of cost to implement machine learning models. Due to faster GPU, the waiting time is less while the code is running.

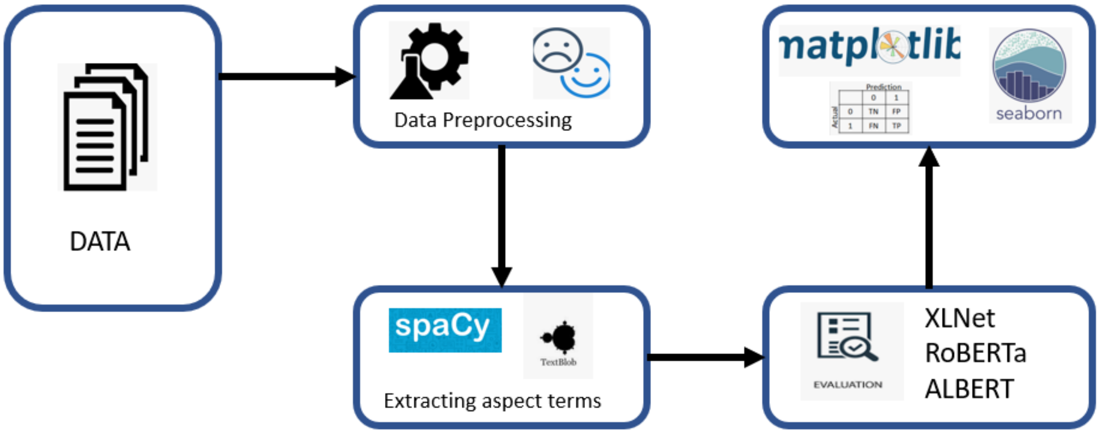

Figure 6: Design Specification

Also, huge setup does not needs to installed in your system for the execution of project and sharing notebooks is very easy. End to End implementation of project can be performed on google colab platform.

5.2 Data Preparation

The dataset obtained after the cleaning the data was of more than 1500000 records. Due to computational limit, 40000 data is considered for modelling. This data is split into train and test set in70:30 ratio respectively. On sentiments data, class imbalance was observed as we can see in figure 4. Table 1 illustrates size of train and test dataset.

	Total Size	Proportion
Training Set	28000	0.70
Testing Set	12000	0.30
Total	40000	1

Table 1: Data Structure

5.3 XLNET

XLNet is a generalized autoregressive pre-training method that leverages the best of NLP task while avoiding their limitations Yang et al. (2019). It is the language representation pre training method developed by CMU and google researchers in 2019. For autoregressive modelling, it uses permutation language modelling. To use the combined language modelling, it uses integrated transformer xl architecture. At several places XL-Net performs better than BERT. BERT generally masks 15% of the tokens. And other important point of word architecture is that it learns this tokens parallelly. XLNet is not just autoregressive model going backward or forward but it works with permutations on both the sides. It does not used LSTM but uses self-attention only. It integrates methods from transformerXL mostly to catch dependencies from far away. Factorization order is one of the mechanisms of the XLNet to predict the order. It is required as it not simply going forward or backward. The sinusoidal signal is just like vanilla transformer. Sequence to train the model is decided by the factorization order N. Random sampling is done to find out all the possibilities. The models need to know the position of the token to predict. So, this features of pre-trained transformed based XLNet technique are used in comparing the aspect based sentiment analysis.

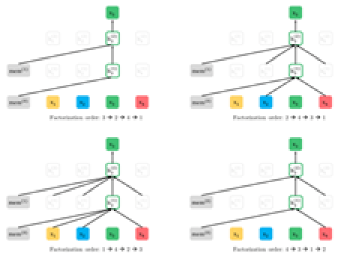

Figure 7: XLNet Factorization referred from Yang et al. (2019)

5.4 ALBERT

A Lite BERT (ALBERT) substantially has few parameters than a traditional BERT (Lan *et al.*, 2019) architecture. ALBERT uses 3 things which makes it better in performance then BERT, which are Factorized embedding Parameterization, Cross layer parameter sharing and Inter sentence Coherence loss. In BERT, the hidden layer size is tied by embedding dimension. It becomes more difficult when the size of hidden layer increases since it increases the embedding dimensions and therefore increases the parameters. To improve the efficiency of parameters, all the parameters across the layers are shared by ALBERT. BERT is trained along with MLM, and it is said NSP task is easy. ALBERT uses the task on which the model has to predict if the sentences are coherent. Therefore, for multi sentence encoding ALBERT models constantly develops downstream (Lan *et al.*, 2019) job performance. Hence, based on these features ALBERT model is used for comparing the performance of aspect based sentiment analysis.

5.5 RoBERTa

RoBERTa is for robustly optimized BERT approach which is configured for improving the end task performance of the BERT model and estimate its collective effect. It has also shown its effectiveness on General Language Understanding Evaluation. The training mechanism is different of RoBERTa compared to BERT. Next Line prediction was used by BERT. Instead of static masking, it is trained with dynamic masking. It was trained on large dataset of English Language text. To extract features both from text and aspect terms, RoBERTa based bidirectional transformer is employed so to treat aspect terms as a subtask. To focus on the most relevant aspect terms and to guide a model, a cross sectional mechanism is applied. The output from the final layer from the input raw text to model is used for feature representation. So, RoBERTa technique is used to compare the aspect based sentiment analysis in this research.

6 Evaluation

For precision over accuracy models were evaluated because to apply sentiment analysis on live reviews data, models should be precise in predicting the sentiments.

6.1 Experiment with XLNet

The dataset is huge, and training was taking a bit longer time so, the model is trained on 40000 data. The data is split into train and test set in 70:30 ratio respectively. The loss in training data was 19.23%. Accuracy is the ratio is correctly predicted values to the total number of observations. When the data is symmetric, accuracy is the good measure. The measure of relevant data points is given by precision. The measure of model correctly identifying the True positives is Recall. The testing accuracy for XLNet was 91% and the precision was 87% and recall was 78%. The F1-score was to be 81%.

6.2 Experiment with ALBERT

Accuracy of the model shows how precisely the values can be predicted. Accuracy is the ratio is correctly predicted values to the total number of observations. When the data is symmetric, accuracy is the good measure. In this research the performance of the models is evaluated based on accuracy. The data was splitted in 70:30 ratio. The training accuracy for ALBERT model was 83.06% and the testing accuracy for the model was 83.80%

6.3 Experiment with RoBERTa

In this the model is directly applied on the sentiments labelled from the star ratings. For evaluating RoBERTa model, training and testing accuracy is calculated. The data was splitted into train and test set in 70:30 ratio respectively. The training accuracy is 92.45% and the testing accuracy is 94%. The below figure is the confusion matrix on which x-axis is the predicted label and on the y-axis is the Actual label. From the figure 8 it can be seen that the value was negative and it predicted negative for 97% and the value was positive and it was predicted positive 76% of the time.

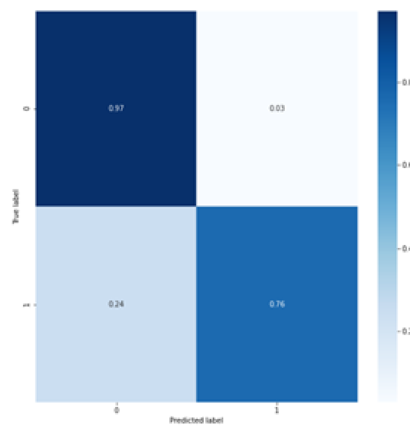

Figure 8: Confusion matrix of RoBERTa

6.4 Experiment with XLNet based on Aspect terms

By using the extracted aspect average score, the model is trained. Since the dataset is huge and training was taking a bit longer time, the model is trained on 40000 data. The data is split into train and test set in 70:30 ratio respectively. Accuracy is the ratio is correctly predicted values to the total number of observations. When the data is symmetric, accuracy is the good measure. The measure of relevant data points is given by precision. The measure of model correctly identifying the True positives is Recall. The testing accuracy for XLNet was 80% and the precision was 80% and recall was 79%. The F1-score was to be 79%.

6.5 Experiment with ALBERT based on Aspect terms

The data was splitted into train and test set in 70:30 ratio respectively.The model was running for 5 epochs on the data. The accuracy for training data is 59.67% and training loss is 0.002 and the testing accuracy is 59.70%. The accuracy of the ALBERT with aspect based method is observed to be very low compared to the accuracy of model with aspect based method.

6.6 Experiment with RoBERTa based on Aspect terms

The training accuracy of the RoBERTa model is 89.06% and also the testing accuracy was the same. The model was run for 2 epochs on the data. As the number of epochs increases the accuracy of training and validation also increases and thus the losses decrease. The below figure is the confusion matrix on which x-axis is the predicted label and on the y-axis is the Actual label. From the figure 9 it can be seen that the value was negative and it predicted negative for 86% and the value was positive and it was predicted positive 90% of the time.
In figure 9, 10 and 11 comparison of models accuracy with and without aspect based approach is shown. In which it can be seen that the accuracy of without aspect based approach models is higher.

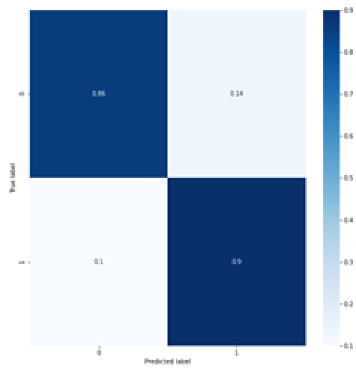

Figure 9: Confusion matrix of RoBERTa with aspect terms

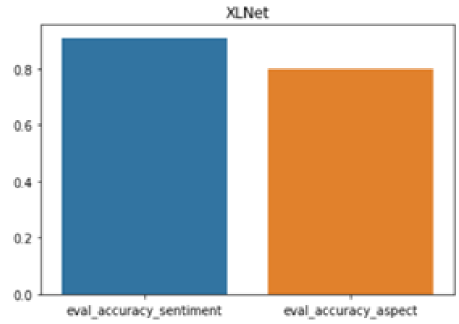

Figure 10: Comparison of XLNet with and without aspects

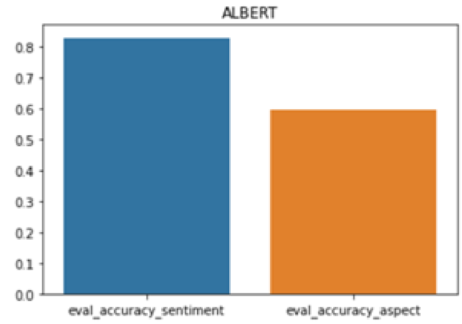

Figure 11: Comparison of ALBERT with and without aspects

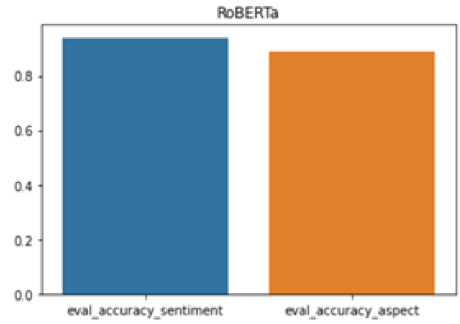

Figure 12: Comparison of RoBERTa with and without aspects

7 Discussion

The intention of the research was to explore the transformer based models with aspect terms. Unlike the traditional approach of sentiment analysis, classifying the review as positive or negative, this approach included identifying the aspect terms and the associated sentiments with it and then taking the average of the sentiments if there were 1 or more sentiments associated.
Major challenges faced were:
Size of the data: The size of the data is too large and due to computational limits; it was not possible to train model on entire dataset since it was taking very high time to train the model, so only 40000 data was used.
Class Imbalance on sentiments data: As the model was to be applied on data with sentiment column and aspect average score data, it was seen there was class imbalance with sentiment data since the number of positive sentiments was very high then the negative sentiments
In total 6 transformer based models were applied and evaluated. It can be said that due to class imbalance with the sentiment data, the accuracy of that data was biased. To improve the performance of models various experiments were carried out. All the model except the ALBERT with aspect based method achieved a precision of more than 80%. In the aspect based approach the highest accuracy achieved was by RoBERTa model which was 89.06% followed by XLNet and ALBERT. The approach without aspect terms achieved greater accuracy and the highest accuracy in that was achieved by RoBERTa model followed by XLNet and ALBERT. From the results it is seen that the XLNet,

RoBERTa and ALBERT models without the application of ABSA has outperformed all other models. The total number of negative labelled sentiments were predicted correctly was higher in the case of RoBERTa without ABSA compared to the one with ABSA approach and the number of positive sentiments predicted were higher in the case of RoBERTa with ABSA approach. XLNet and RoBERTa both with ABSA approach has achieved a good amount of accuracy but was not good enough to outperform models without ABSA technique. From the results , it can be said that the addition of aspects has not affected the model performance to a great extent but making state of the art model has performed better without ABSA approach. The number of negatively labelled data was less due to which there was class imbalance. Additionally, the performance of the model is affected by the imbalance nature of data significantly.

8 Conclusion and Future Work

This section summarizes the whole research and presents the answer to the research question. Also, the impact of this research along with the future work is discussed which can carried out by other researchers. The main focus of the research is to improve the performance of transformer based models by applying aspect based sentiment analysis. The whole research was broken into 3 parts: selecting and pre processing the data, extracting the aspect terms and applying the models with and without aspect based method. The dataset used for this research was downloaded from amazon web services official website. Furthermore, the extraction of aspect terms was done by using spaCy library. After extracting the aspect terms and sentiments associated, the resulting average was stored in one of the columns (aspectAverageScore) and was added to the dataset for implementation. Then the models were executed on both with and without aspect based approach and was evaluated on the basis of accuracy. RoBERTa has outperformed in both the scenarios with and without ABSA technique. The model had performed better on data without aspect terms. These results helped to answer the research question: Can Aspect Based Sentiment Analysis (ABSA) technique with the obtained aspect terms improve the performance of pre-trained models in comparison to the same models without the ABSA technique for Sentiment Classification using Amazon data? The answer is: no, it cannot. For future work, more data should be considered by increasing the computational limits. Also, the class imbalance needs to be handled. By handling the class imbalance, the performance of aspect based approach can be better. To verify the robustness, the framework should be applied to different domain dependent datasets and for verify the performance of pre-trained models, more variants of BERT and other pre-trained models should be applied.

Acknowledgement

I would like to convey my gratitude to my supervisor Dr. Bharathi Chakravarthi who has constantly extended required support throughout the research work including the guidance on standards of academic research documentation. His continuous feedback helped me to present this research project in an insightful and sophisticated manner.

References

Adoma, A. F., Henry, N. M. and Chen, W. (2020). Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition, pp. 117–121.

Alshahrani, A., Ghaffari, M., Amirizirtol, K. and Liu, X. (2020). Identifying optimism and pessimism in twitter messages using xlnet and deep consensus, pp. 1–8.

Basiri, M. E. and Kabiri, A. (2018). Uninorm operators for sentence-level score aggregation in sentiment analysis, *2018 4th International Conference on Web Research (ICWR)*, pp. 97–102.

C.P., H. and Sunitha, N. (2020). Topic identification for semantic grouping based on hidden markov model, *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 932–937.

Do, H. H., Prasad, P., Maag, A. and AbeerAlsadoon (2019). Deep learning for aspect-based sentiment analysis: A comparative review, *Expert Systems with Applications* **118**: 272–299.
URL: *https://www.sciencedirect.com/science/article/pii/S0957417418306456*

Duan, J., Zhao, H., Zhou, Q., Qiu, M. and Liu, M. (2020). A study of pre-trained language models in natural language processing, *2020 IEEE International Conference on Smart Cloud (SmartCloud)*, pp. 116–121.

Faguo, Z., Fan, Z., Bingru, Y. and Xingang, Y. (2010). Research on short text classification algorithm based on statistics and rules, *2010 Third International Symposium on Electronic Commerce and Security*, pp. 3–7.

Ghasiya, P. and Okamura, K. (2021). Investigating covid-19 news across four nations: A topic modeling and sentiment analysis approach, *IEEE Access* **9**: 36645–36656.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. and Soricut, R. (2020). Albert: A lite bert for self-supervised learning of language representations, *ArXiv* **abs/1909.11942**.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach, *ArXiv* **abs/1907.11692**.

Ma, Y., Li, Y., Wu, X. and Zhang, X. (2018). Chinese text classification review, *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 737–739.

Myagmar, B., Li, J. and Kimura, S. (2019). Cross-domain sentiment classification with bidirectional contextualized transformer language models, *IEEE Access* **7**: 163219–163230.

Pham, Q. H., Anh Nguyen, V., Doan, L. B., Tran, N. N. and Thanh, T. M. (2020). From universal language model to downstream task: Improving roberta-based vietnamese hate speech detection, *2020 12th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 37–42.

Pipalia, K., Bhadja, R. and Shukla, M. (2020). Comparative analysis of different transformer based architectures used in sentiment analysis, *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*, pp. 411–415.

Shahnawaz and Astya, P. (2017). Sentiment analysis: Approaches and open issues, *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 154–158.

Shirsat, V. S., Jagdale, R. S. and Deshmukh, S. N. (2017). Document level sentiment analysis from news articles, *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, pp. 1–4.

Singh, A. K. and Verma, A. (2021). An efficient method for aspect based sentiment analysis using spacy and vader, *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 130–135.

Soni, P. K. and Rambola, R. (2021). Deep learning, wordnet, and spacy based hybrid method for detection of implicit aspects for sentiment analysis, *2021 International Conference on Intelligent Technologies (CONIT)*, pp. 1–6.

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). Attention is all you need, *ArXiv* **abs/1706.03762**.

Wang, H., Hu, X. and Zhang, H. (2020). Sentiment analysis of commodity reviews based on ALBERT-LSTM, *Journal of Physics Conference Series*, Vol. 1651 of *Journal of Physics Conference Series*, p. 012022.

Wang, Y., Chen, Q., Ahmed, M., Li, Z., Pan, W. and Liu, H. (2021). Joint inference for aspect-level sentiment analysis by deep neural networks and linguistic hints, *IEEE Transactions on Knowledge and Data Engineering* **33**(5): 2002–2014.

Yang, J. and Yang, J. (2020). Aspect based sentiment analysis with self-attention and gated convolutional networks, *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 146–149.

Yang, X., Yan, B., Li, H. and Chen, Y. (2020). Retransformer: Reram-based processing-in-memory architecture for transformer acceleration, *2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, pp. 1–9.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. and Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding, *NeurIPS*.

Yao, T., Zhai, Z. and Gao, B. (2020). Text classification model based on fasttext, *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)*, pp. 154–157.

Zhai, Y., Song, W., Liu, X., Liu, L. and Zhao, X. (2018). A chi-square statistics based feature selection method in text classification, *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 160–163.

Zhang, L. and Liu, B. (2017). Sentiment analysis and opinion mining, *Encyclopedia of Machine Learning and Data Mining*.

Zhang, T., Xu, B., Thung, F., Haryono, S. A., Lo, D. and Jiang, L. (2020). Sentiment analysis for software engineering: How far can pre-trained transformer models go?, *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pp. 70–80.

Zhang, Y. and Rao, Z. (2020). n-bilstm: Bilstm with n-gram features for text classification, *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pp. 1056–1059.

Zhao, N., Gao, H., Wen, X. and Li, H. (2021). Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis, *IEEE Access* **9**: 15561–15569.