National College of Ireland

Utilizing the Transformer models for Analysing Deceptive Reviews and Aspects of the reviews

MSc Research Project
MSc Data Analytics

Santhosh Vinayagamurthy
Student ID: x20186177

School of Computing
National College of Ireland

Supervisor: Prasanth Nayak

National College of Ireland

MSc Project Submission Sheet

Student Name: Santhosh Vinayagamuthy…….

Student ID: …X20186177……..……...

Programme ……MSc Data Analytics…. **Year:** …September 2021…

Module: …Msc Research Project……

Supervisor: Prashanth Nayak……..……..

Submission Due Date: 15/12/2022……

Project Title: … Utilizing the Transformer models for Analysing Deceptive Reviews Aspects of the reviews

Word Count: ……6700…… **Page count**…18……..……..

School of Computing

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: ……Santhosh V……

Date: ……15.12.2022……

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):	

Utilizing the Transformer models for Analysing Deceptive Reviews and Aspects of the Reviews

Santhosh Vinayagamurthy
x20186177

Abstract

The customer posts their reviews of a product on social media platforms and on e-commerce platforms about a product. These reviews can be read by other customers who are also willing to buy the product. Getting an idea about a product before buying it is good for the customers. But some people or organizations in order to promote their products start posting fake positive reviews. Similarly, to demote their opponent's business they post fake negative reviews. This can misguide the customer's purchase decisions. Several researchers previously addressed this issue using machine learning and deep learning models. The transformer model like BERT, RoBERTa, and DeBERTa are used. A novel classifier DeBERTa has been explored in this research. The results show that the DeBERTa model outperforms the BERT and ROBERTA models and has achieved an F1 score of 93%

Contents

1 Introduction

1.1 Background & Motivation

The customers have been used to refer to the reviews posted by other customers before purchasing any products on the e-commerce platforms. Referring to reviews posted by other customers gives new customers an idea about a product they intend to buy. The report shows the decision to buy a product or not buy a product depends on how many positive reviews and negative reviews a product has. Some of the spammers make use of this and started to spam their opinion in favor of the sellers to gain some monetary benefits from the sellers. This is referred to as Opinion Spamming(Hai et al., 2016).

This has become a social concern now as customer mindsets can be changed based the fake reviews. These actions by the spammers can be misleading customers by promoting or demoting products unfairly to the customers. The motivation behind these fake reviews could be, the manufacturers can pay the reviewers to promote a product to the customers even though the quality of the product was not good. The other scenario can be the sellers or marketers who would want to deliberately bring down the reputation of their opponent's market. Also, a report shows that the percentage of fake reviews has increased from 5% in 2006 to 20% in 2013(Hai et al., 2016). Considering all these points makes this an important issue that has to be analyzed(Kim et al., 2021) further with the latest advancements.

Fake reviews have been first analysed by (Jindal & Liu, 2008). Deceptive review analysis has been considered a classification problem. To detect fake reviews logistic regression has been used. Similarly, several machine learning and deep learning algorithm have been used for analyzing fake reviews(Elmurngi and Gherbi, 2017). Several machine learning models like Naïve Bayes(NB), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) were used to find fake reviews. The SVM algorithm outperforms all the models in analyzing fake reviews. The drawback of using machine learning models for this analysis is that extracting the features is difficult. To improve accuracy deep learning models have been used for detecting fake reviews.

Some of the deep learning models like Long short-term memory (LSTM), Convolutional Neural Networks (CNN), and Recurrent Neural Network(RNN) has been used by several researchers for analysis of fake reviews. These deep learning models can be used directly or can be ensembled with each other. The CNN-LSTM is a model proposed by (Alsubari et al., 2021), this model has been applied to the different standard fake review datasets for analysing the fake reviews in the in-domain and cross-domain. The LSTM model is combined with CNN to analyze the contextual information from the texts. The algorithms like CNNLSTM-FABC have been used by (Jacob and Selvi Rajendran, 2022) to detect deceptive reviews. Using CNN and LSTM, Fuzzy artificial bee colony (FABC) a hybrid model has been introduced , and

performs better than other hybrid models. The drawback of using deep learning models is they can be used for large datasets and the doesn't provide any computation parallelly.
To overcome this disadvantage the transformer models have been introduced. Transformer models are also considered pre-trained models which are already with a huge dataset and also work in analysing the texts in both directions. Some of the transformer models like BERT. RoBERTa and DISTILBERT have been previously used in analyzing fake reviews. Among them, the RoBERTa model performed well(Gupta et al., 2021).

Several natural language processing models have been used by researchers to analyse text and fake reviews. These models can be used for analyzing the sentiments present in the texts and also can be used for classification, this is termed sentiment analysis. Sentiment analysis falls at the intersection of Information retrieval, Natural language processing, and machine learning. It is the process of analysing the sentiments in the text and finding whether the given text or sentence is positive, negative, or neutral. Similarly, the process of analyzing the aspects present in the text or sentence is called aspect sentiment analysis(Mowlaei et al., 2020). This research deals DeBERTa model (novel transformer model) has been explored in for analysing the fake reviews and aspects present in it. The Disentangled mechanism in the DeBERTa model has helped it to perform better than the BERT and RoBERTa models in some benchmark datasets(He et al., 2020).

Figure 1: Aspect-Based Sentiment analysis.

1.2 Research Question & Objective

1.2.1 Research Question

What degree of effectiveness does the DeBERTa model provides in improving the classification of fake reviews, and analyzing the aspects present in it when compared to BERT and RoBERTa models.

1.2.2 Research Objective & Contribution.

In this research, the DeBERTa model which is not yet been explored.

- Data Pre-processing has been done.
- The F1 score and the accuracy of analysing the fake reviews have been compared between the BERT, RoBERTa, and DeBERTa models.
- The DeBERTa model has been used for analyzing the aspects present in the reviews.

1.3 Structure of the Paper

The structure of this research report is as follows. The related works section follows, and it offers information on earlier studies that were carried out to analyze fraudulent reviews. The Methodology section consists of the project overviews section which briefs about the techniques and data used. The project's implementation details are discussed in detail in the

next part, which is followed by the results section. The conclusion and future work are described in the end.

2 Related Work

This section of the paper discusses the literature reviews on detecting fake reviews in chronological order. This section gives an overview of how deceptive reviews are detected using different techniques (Machine Learning, Deep Learning, and Transformer models) and addresses the limitation and pros present in those papers.

2.1 Detecting Fake Reviews using Machine learning (ML) and Deep learning(DL)

The first fake review detection was done (Jindal and Liu, 2008) with machine learning models like naïve bayesian algorithm, SVM, and Logistic regression has been used. Out of these algorithms, Logistic regression performed better. The limitation here in this research was it was very hard to distinguish between fake and real reviews, so they considered the duplicates as fake. Similarly, several machine learning algorithms have been used in fake reviews analysis. (Hassan and Islam, 2020) have used ML models like logistic regression, SVM, and Naïve Bayes classifier among these the SVM classifier has given an accuracy of 88.75%. The drawback here is small dataset had been for the analysis.

The Machine Learning classifier algorithms are used for detecting fake reviews. Models like SVM, Naïve Bayes, Decision tree, and Logistic regression have been used among them the SVM performed better than the other models. The limitation here is dataset is imbalanced and TF IDF is the only feature extraction technique used (Khan et al., 2021).

To find fake reviews ML classifiers like BAE(Bootstrap aggregate ensembles), K-Nearest Neighbour KNN, random forest, neural network, and logistic regression have been used by (Lee et al., 2022)Linguistic Inquiry and Word Count (LIWC) is the pre-processing technique for extracting linguistic features. The dataset used in this experiment belongs to yelp so the performance of the classifier decreases other domains or sites. Taking into account these negative aspects like not being able to analyze different domain data and extracting the features from machine learning algorithms, researchers started to use deep learning models to find fake reviews. Several researchers use the models like LSTM(Long-short term memory), CNN(Convolutional Neural Network), and RNN (Recurrent Neural Network) for finding deceptive reviews.

A novel approach called hierarchical attention architecture has been introduced to find the deceptive reviews. It has two components combined together one is the convolutional structure and the other is the Bi-LSTM (Bidirectional Long-short term memory). The convolution network works along with word2sent-level to extract the information. The Bi-LSTM and Sent2Doc-level are used for context extraction. The proposed model performed better than the machine learning models. The drawback here is the model is trained with labelled data so predicting the unable data is not possible (Liu et al., 2022).

The hybrid model which is a combination of CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network) was built to improve the accuracy of fake review detection. This model helps to capture the sequential information in the text. Two different datasets have been used in this analysis FA-KES and ISOT also the hybrid model is compared with several machine learning algorithms like linear regression, KNN, CNN, and RNN, a Decision Tree among these algorithms the CNN-RNN performed better (Nasir et al., 2021).

2.2 Detecting Fake Reviews using Transformer models.

The deep learning model needs a huge dataset for training the models and also requires more computational power. Understanding the meaning of the text in these models is hard. To overcome these the transformer models have been introduced. The Transformer models can also be used for the dominant sequence transduction problems. This model contains an additional feature encoder and decoder along with an attention feature. Two translation tasks have been done using the transformer model techniques which have produced a better BLEU of 28.4 which is 2 BLEU (Bilingual Evaluation Understudy) more than the other deep learning models. The features like parallelization and less time consumption make this model very powerful and the fixed length content is considered a disadvantage here (Vaswani et al., 2017).

With the base of the transformer models, several pre-trained models have also been developed like BERT (Bidirectional encoder Representation). Using the BERT model, the content present in the text can be analyzed in both directions. As these models are already pretrained one output layer can be added at the end to achieve good results in the natural language processing tasks. With the help of the Masked Language Model (MLM), bi-directionality is achieved in analyzing the texts (Devlin et al., 2019). Several researchers have used the BERT model for classification problems and finding the text's polarity. For finding the polarity of the reviews in the IMDB dataset (Abdul et al., 2019) used the BERT model. This model has achieved an F1 score of 89% in finding the polarity in the reviews(Abdul et al., 2019).

The length of the reviews may vary. In some of the reviews, the reviews may be shorter in some the reviews are longer. Extracting the feature from the shorter text is difficult. To extract the feature (Hu et al., 2022) have proposed research using the BERT model. The mental features of the reviewers have been analyzed to find the feature from the short texts. Several machine-learning algorithms like Support vector machines and deep learning models like concurrent neural networks and recurrent neural networks have been used for comparison with the BERT models. The usage of the mental features of the reviewers improves the accuracy of predicting fake reviews with short texts. BERT does a better job of combining the mental features with the short texts when compared to other models. The drawback of this method is the inference factor can change based on different domains.

Similarly, the classification of the helpful and unhelpful reviews is done using the transformer models. Different bag of words classifier like SVM, KNN, and NB has been compared with the BERT model in this experiment The TFIDF method is used for extracting the feature from the text after that the input is passed to the machine learning models. The results show that the BERT model performed better in analyzing fake reviews. The dataset used is extracted from yelp reviews. The limitation of this approach is that the dataset used does not contain the varying length of the text and other BERT-based models are not explored.

To analyze fake reviews (Kim et al., 2021) have proposed a novel approach called YONG (You only need Gold) which is an assisting tool along with the reviewer's behavior. The following are considered as the indicator for finding the fake reviews model decision, probability, and evidence. The output predicted by the model whether the given review is fake or not is called the model decision. Probability indicates the SoftMax value which shows how confident the model is in predicting the values. The Evidence indicates based on what factor/word the reviews are classified as fake or real. Several comparisons of the F1_score and the accuracy of the different models have been done like BERT and other conventional machine learning algorithms SVM, FFN, CNN, and LSTM. Also, the performance of the BERT model based on the with and without indicators has been done. The BERT model performed better when compared to conventional models and similarly, the BERT produced good results when the tools are used for the prediction.

To identify fake news (Karande et al., 2021) have used several deep-learning models like LSTM, Bi-LSTM, and CNN with different embedding models. First glove embedding, glove embedding with attention mechanism, and BERT embedding has been used. Among them, the BERT embedding has performed better than the other models. Similarly, the BERT classifier has also been used for analyzing the fake reviews it performed better than the machine learning models.

RoBERTa is also a pre-trained model which has been built on the basis of transformer models and BERT. (Liu et al., 2019) considers that BERT models are less trained and the amount of pre-trained data which are used in the BERT model can be increased to get more accurate results. The RoBERTa model performed better than the BERT model and the best result are achieved in the SQUAD and GLUE datasets with less finetuning.

The new architectural model to boost the Transformer models' performance. The DeBERTa model outperforms the BERT and RoBERTa models because it utilizes the disentangled attention technique. Each word is characterized as a vector, where both its content and position are encoded. The weights for the content and positioning of the words are calculated using the disentangled matrices. Additionally, it has an improved masked decoder. An improved adversarial strategy is offered to help fine-tune the model and will result in better outcomes. When applied to the SuperGlue Dataset, the DeBERTa model outperformed the BERT and RoBERTa models (He et al., 2020).

2.3 Detecting Fake Reviews using BERT, ROBERTA, ALBERT (Multiple transformer models).

For classifying Covid-19 fake news on Twitter (Qasim et al., 2022) used several transformer models. In this experiment transformer models like BERT-base, BERT-large, RoBERTa-base, RoBERTa-large, DistilBERT, XLM-RoBERTa-base, ALBERT-base-v2, Electra-small, BART-large have been used. Two different datasets have been used in this experiment COVID-19 fake news dataset and COVID-19 English tweet dataset, and the extremist-non-extremist dataset. For the fake news covid-19, RoBERTa-large has performed better than the other model. For the Covid-19 English tweet dataset, Bart-large performed better.
The transformer models have also been used for analyzing fake reviews. The transfer models like BERT are considered the state-of-the-art approach for analyzing fake reviews. The dataset used in this experiment is collected from yelp reviews. The distribution of the values dataset between fake and original reviews is 60% and 40 %. Several transformer models like BERT, RoBERTa, DistillBERTa, and AlBERT have been used here. Before fine-tuning the reviews

into the model the pre-processing steps are followed to remove the punctuation, remove the URL, and so on. The results show that the RoBERTa model has achieved good accuracy in detection of the fake reviews than the other transformer models (Gupta et al., 2021).

2.4 Detecting Aspects in Fake Reviews

To find the aspects present in the text (Hoang et al., n.d.) used the BERT model. To find the text aspect which does not belong to the same domain the BERT model is fine-tuned with additional text and sentence pair classification. The out-of-domain aspect analysis is done on the SemEval dataset. Three types of aspect analysis have been done in this one aspect category classifier, sentiment classifier, and both of them combined. The aspect classifier is trained with labeled data related and unrelated. The relation between the aspect and reviews is used for the sentiment classifier. Both the sentiment are fed into the combined classifier for analyzing the aspect. The combined classifier performed better than other models and achieved better scores.

Customer review sentiment and aspects have been examined using four different models. The Naive Bayes method, BERT, Support vector machine, and LSTM are the models used in the research (Geetha and Karthika Renuka, 2021). The feature is the additional part that is needed to find the sentiment and aspects present in the reviews using the Naïve Bayes, SVM models in the small dataset. LSTM performs better than the traditional algorithms but cannot be considered fully bidirectional. Being able to analyze the text in both ways sets the BERT model apart from the other models. This enables the BERT model to generate cutting-edge text classification outcomes. Before fine-tuning, a number of pre-processing procedures such as canonicalization and tokenization were performed.

In order to demonstrate how the detachment of the position and content vectors can benefit the performance of the ABSA tasks, this research uses a disentangled attention mechanism for sentiment analysis. Position and content are different from one another in a somewhat similar way to how syntactic and semantic spaces are separated, with position referring solely to the text's ordering feature and leaving other syntax features. The ABSA-DEBERTA achieved an F1 score of 81.39 when compared to other models like BERT-SPC, LCF-BERT, and BAT.

3 Research Methodology

This research uses the Knowledge discovery Databases (KDD) methodology. This helps in identifying trends in data from large-volume datasets. The methods, data, and approaches selected for this research on deceptive reviews are covered in detail in the section that follows.

3.1 Project Overview

- The pre-trained models based on transformers are used to analyze fraudulent reviews. For this research, models like BERT, RoBERTa, DeBERTa, and have been picked.

- Microsoft has proposed a brand-new architecture called DeBERTa. It was constructed with BERT architecture.

- DeBERTa (Decoding-enhanced BERT with Disentangled Attention) model outperforms other transformer models because it includes features like disentangled attention and an advanced masked decoder.

- The dataset that has been selected for this proposal will be divided into a training set and a test set.

- To detect the misguided reviews and real reviews in the dataset and analyze the aspects existing in the dataset, the Transformer models will be trained or fine-tuned using the training set.
- Using evaluation measures like F1 score, recall, and precision, the effectiveness of the transformer models BERT, RoBERTa, DeBERTa, and AlBERT must be compared.

3.2 Data Understanding

The dataset used for this project consists of 2.5k deceptive reviews and 2.5K original reviews. This dataset is the publicly available dataset. The original reviews are fetched from amazon customer reviews. The deceptive reviews are created with the help of models such as GPT and ULFIT. Most recently this dataset has been created and has not been analyzed or used in any of the research to the best of my knowledge(Salminen et al., 2022).

3.3 Data Processing

Data pre-processing is done initially before feeding the data into the models. The elimination of null and duplicate values has been carried out. Characters like punctuation, numbers, stop words removal and URLs have been removed. The processing of tokenization was performed by the models applied in the dataset as it contains the inbuilt tokenizer in it.

3.4 Data Modelling

The BERT model is built based on the transformer model's feature like parallelization and bidirectional making the BERT model better than the deep learning models and machine models This research exploits how the novel pre-trained model DeBERTa, can be used for analysing the fake reviews. The characteristic of the DeBERTa model like the disentangled attention mechanism has made it more powerful than the BERT and RoBERTa model. This DeBERTa model has been used for analysing the aspects and its performance has been compared with BERT and RoBERTa.

4 Design Specification

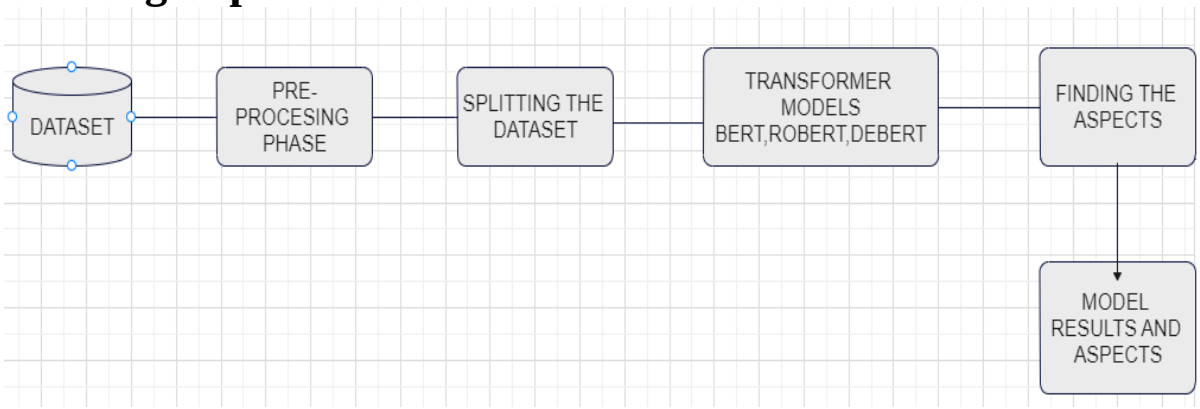

Figure 2: Model Architecture.

The details of this architecture have been explained in section 5. This section covers the details of the Transformer models like BERT, RoBERTa, and DeBERTa.

4.1 Transformer model.

Since the deep learning models struggle in parallelization to overcome this issue the transformer models are utilized. The transformer model performs sequence transduction by establishing a connection between the encoder and the decoder via the attention mechanism. Numerous translation outcomes demonstrate that these models allow parallelization and produce good results despite having had such a short training period. The transformer model has 6 encoders and 6 decoders. A feed-forward neural network and a self-attention layer are present in each encoder. A third layer known as encoder-decoder attention is included in the decoder along with the same other two layers like the encoder. To convert the word into a vector the word embedding layer has been utilized. This embedded value will be passed up till the last encoder. This model can also incorporate multiple-head self-atte9ntion, which can enhance NLU task performance. In order to determine the following predicted probabilities, the obtained self-attention matrices will be supplied to the decoder output to the SoftMax. The transformer model performs better than the other models for translation tasks(Vaswani et al., 2017).

4.2 BERT

Based on the transformer's encoder architecture, the BERT (Bidirectional encoder Representation from Transformer) model was created This model is bidirectional, and pre-trained from the unlabelled text. Fine-tuning the BERT model with an output layer can give the best results. BERT has the capacity to comprehend the context of the words in a bidirectional manner with the aid of the masked language mode. The BERT model has enhanced both the task at the sentence level and the token level. The BERT model tokenizes the words using word piece tokens. There are two main types of BERT, BERT Base, and BERT Large. The BERT Base has 12 Transformer blocks (L), a hidden size of 768 (H), and 12 attention heads (A) with a total of 110M parameters. BERT Large (24 Transformer blocks(L), 1024 hidden size(H), and 16 total parameters(A)) has 340M parameters in total. The BERT model is assisted in comprehending the context of the sentence by features like Masked language modeling, Next sentence prediction, and pre-trained nature(Devlin et al., 2019; Vaswani et al., 2017). Since the BERT model contains these types of features so this model has been used in this research.

4.3 RoBERTA

According to (Liu et al., 2019),the BERT model is undertrained, and an upgraded model termed the RoBERTa model has been developed (Robustly optimized BERT). This model outperforms the BERT model because it was pre-trained effectively using better pre-training procedures. When compared to BERT, the RoBERTa model now has four additional modifications. A huge amount of information is used to train the RoBERTa model for more time. The long series of data was used for its training. The Next sentence prediction feature is no longer available. The BERT model uses a single static mask.Training data was duplicated ten times, resulting in being masked ten different ways throughout the training of 40 epochs. This was done to prevent utilizing the same mask for each and every training instance in every epoch. So to overcome this the dynamic changing mask pattern has been used.The RoBERTa was developed using the BERT model and these four improvements produced good results in GLUE and MNLI datasets. A total of 160 GB of data is used for the pre-training of the RoBERTa model, which uses five different corpora from diverse streams. It used byte-pair encoding for tokenization for dynamically masking the tokens. In light of this, the RoBERTa model will be used in this study to categorize opinion spamming

4.4 DeBERTa

Decoding-enhanced BERT with disentangled attention is a powerful new strategy that has been put out (He et al., 2020). On Considering the tasks involving natural language processing, this model outperforms the RoBERTa and BERT model. To enhance the performance of this model in text analysis, two new features have been introduced. They have improved the mask decoder and disentangled the attention mechanism. A brand-new adversarial tuning technique has been implemented to fine-tune the model.

Disentangled attention

DeBERTa is pre-trained utilizing masked language modeling(MLM)., the same as BERT. MLM is a fill-in-the-blank challenge where a model is trained to predict the masked word from the words around the token. A single vector will be utilized in the BERT model to capture both the location and content of a word. In contrast, the DeBERTa model represents the position and the content using two vectors. To get the weight for these vectors, the disentangled matrix is used. This demonstrates how the word's position and meaning are taken into account when determining the attention weight.

Enhanced mask decoder

Prior to the softmax layer, DeBERTa model offers the absolute word position embedding technique to decode the masked word. The position and the word's context-based embedding were utilized for the masking. The absolute position is present in the input layer of the BERT model. The relative position is present in the transformer model for the DeBERTa model. The relative position acts as the important term for decoding the text whereas the absolute position is considered as the additional feature.

Scale-invariant fine-tuning

Adversarial training techniques are applied to fine-tune the model to enhance the generalization. The approach used to normalize the word embeddings and transform them into a stochastic vector is called scale-invariant fine tuning (SiFT). It achieves better results than the other models like BERT and ROBERTA models while performing on the BenchMark dataset, such as SuperGlue, is taken into consideration.

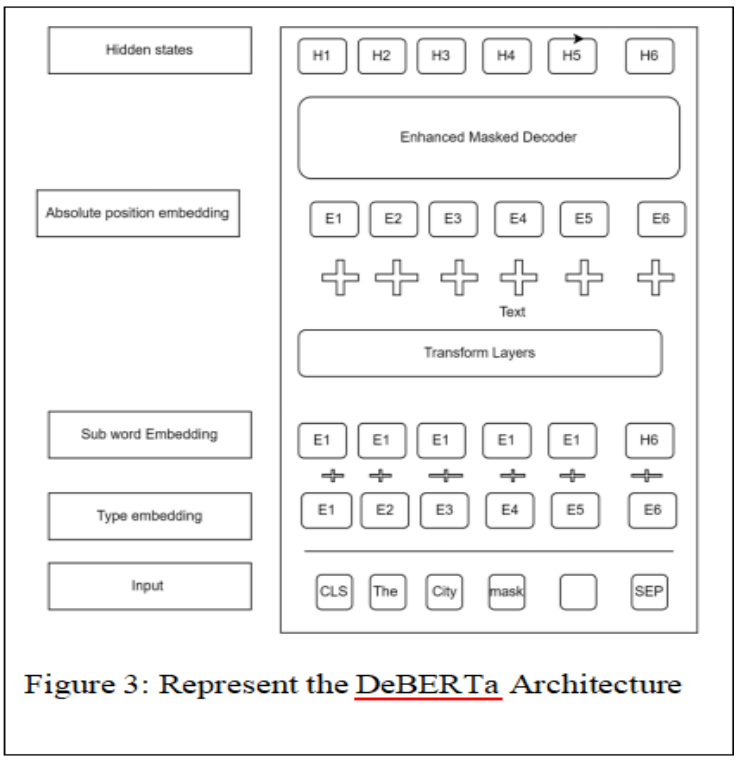

Figure 3: Represent the DeBERTa Architecture

5 Implementation

In this section, the step-by-step approach of how the project is implemented has been explained in detail.

IDE	PROGRAMMING LANGUAGE	FRAMEWORK/LIBRARY	GPU TYPE	Number of GPU
Google Colab(trail version)	Python	SimpleTransformer, HuggingFace Transformer, Sklearn, Pandas, Numpy,PYTorch,PYABSA	Used the trial version GPU Persistence-M	1

Table 1: Configuration

Data processing phase:

The dataset used for this experiment is publicly available. The dataset contains almost equal of amount fake reviews and original reviews. 2.5K fake reviews and 2.5K original reviews. The original reviews are extracted from amazon. The format of the dataset is CSV, which is uploaded in the colab and it is converted as the data frame using panda's library. Initially, the dataset consists of five columns out of those columns only two columns are required for analyzing the fake text column and label column. The extra columns have been dropped. In the next step, the label column contains information about the reviews whether it is fake or original. This column has been encoded and categorized as 0 and 1. 0 represents the fake reviews and 1 represents the original reviews. Then the dataset has been split into the train set and the test set. In the 70:30 ratio, data-cleaning activities like duplication removal, URL removal, and null values removal have been performed this dataset does not contain any null values.

Once the data pre-processing and the splitting is the done the data is now loaded into the transformer models. The training set has been passed initially to the models like BERT, ROBERTA and, DEBERTA. The simple transformer model has been used for this research. This model takes the hugging face library as the base and runs on top of it. Using the hugging face library the pre-trained model like BERT, ROBERTA and DEBERTA API can be called and used for several NLP tasks. The simple transformer makes it easy to access these APIs and fine-tune the model with required data and perform the NLP task.In this research project, the simple transformer is used to get the transformer model the model has been fine-tuned with the review spam dataset. Different types of transformer models API has been called using the simple transformer to carry out this experiment.

6 Evaluation

This section discusses the performance of the transformer models like BERT, ROBERTA, and DeBERTa for deceptive review classification.

Precision: Less false positives should be produced by an effective deceptive analysis model. The model's dependability will suffer if the false positive rate is large. The precision measures how many fake reviews rightfully belong to the fake reviews.

Recall: The recall score indicates whether all the relevant fake reviews are retrieved. The number of fake reviews that are predicted by a model to the entire count of fake reviews that were predicted. To maintain a good recall score the model should not predict the original review as fake. The recall is therefore regarded as the best and most crucial metric that must be taken into account for identifying the fake reviews model.

F1_Score: The F1_score can be obtained by combining the precision with the recall. The total ability of the deceptive review model can be measured using this.

6.1 Experiment / Case Study 1

The first experiment deals with the result of the BERT model (BERT-base-uncased). The BERT model has an inbuilt tokenizer for tokenizing the text this is called a word-piece tokenizer. The BERT model is trained with different epochs like 3,2,1. The model produces the best F1 score of 93% and an accuracy of 93% in the 3rd epoch. The model was trained with a learning rate of 1e-4 to reduce overfitting. The max sequence length value has been assigned to 128. The Adam epsilon is kept as 1e-8 and the Training batch size is eight. The performance of the BERT started to reduce when we trained with the 4 epochs.

MODEL BERT	Epoch =3	Precision	recall	F1 Score	Support
	0	0.93	0.93	0.93	515
	1	0.93	0.93	0.93	485
	accuracy			0.93	1000
	Macro avg	0.93	0.93	0.93	1000
	Weighted avg	0.93	0.93	0.93	1000

Table 2: BERT WITH EPOCH 3

MODEL BERT	Epoch =2	Precision	recall	F1 Score	Support
	0	0.90	0.94	0.92	519
	1	0.93	0.89	0.91	481
	Accuracy			0.91	1000
	Macro avg	0.91	0.91	0.91	1000
	Weight avg	0.91	0.91	0.91	1000

Table 3: BERT WITH EPOCH 2

MODEL BERT	Epoch =1	Precision	recall	F1 Score	Support
	0	0.93	0.90	0.91	525
	1	0.89	0.92	0.90	475
	accuracy			0.91	1000
	Macro avg	0.91	0.91	0.91	1000
	Weight avg	0.91	0.91	0.91	1000

Table 4: BERT WITH EPOCH 1

MODEL BERT	Epoch =4	Precision	recall	F1 Score	Support
	0	0.89	0.97	0.93	506
	1	0.96	0.88	0.92	494
	accuracy			0.93	1000
	Macro avg	0.93	0.93	0.93	1000
	Weighte avg	0.93	0.93	0.93	1000

Table 5: BERT WITH EPOCH 4

6.2 Experiment / Case Study 2

The RoBERTa model has been used for the second experiment. The Byte-Pair-Encoding tokenizer. The training rate is kept as same as the BERT model as 8. The max sequence rate is 128. The learning rate here is 1e-4. RoBERTa model does not perform well in this dataset. It has achieved the maximum F1 score and accuracy of 68% and 35%.

MODEL RoBERTa	Epoch =3	Precision	recall	F1 Score	Support
	0	0.52	0.1	0.68	516
	1	0.0	0.0	0.00	484
	accuracy			0.52	1000
	Macro avg	0.26	0.50	0.34	1000
	Weight avg	0.27	0.52	0.35	1000

Table 6: RoBERTa WITH EPOCH 3

MODEL RoBERTa	Epoch =2	Precision	recall	F1 Score	Support
	0	0.51	1.0	0.68	511
	1	0.0	0.0	0.00	489
	accuracy			0.51	1000
	Macro avg	0.26	0.50	0.34	1000
	Weighted avg	0.26	0.51	0.35	1000

Table 7: RoBERTa WITH EPOCH 2

MODEL RoBERTa	Epoch =1	Precision	recall	F1 Score	Support
	0	0.50	1.0	0.67	499
	1	0.0	0.0	0.00	501
	accuracy			0.50	1000
	Macro avg	0.25	0.50	0.33	1000

	Weight avg	0.25	0.50	0.33	1000

Table 8: RoBERTa WITH EPOCH 1

6.3 Experiment / Case Study 3

The DeBERTa model uses the Byte-level Byte pair encoding for tokenizing the text. The DeBERTa model performed better with 2 epochs and achieved an F1_score and accuracy of 93%. The Training batch is the same as the other model and similarly, the learning rate and adam epsilon values were been kept as same for all three models.

MODEL DeBERTa	Epoch =3	Precision	recall	F1 Score	Support
	0	0.91	0.92	0.91	522
	1	0.91	0.90	0.90	478
	accuracy			0.91	1000
	Macro avg	0.91	0.91	0.91	1000
	Weight avg	0.91	0.91	0.91	1000

Table 9: DeBERTa WITH EPOCH 3

MODEL DeBERTa	Epoch =2	Precision	recall	F1 Score	Support
	0	0.93	0.94	0.93	531
	1	0.93	0.94	0.92	469
	accuracy			0.93	1000
	Macro avg	0.93	0.93	0.93	1000
	Weight avg	0.93	0.93	0.93	1000

Table 10: DeBERTa WITH EPOCH 2

MODEL DeBERTa	Epoch =1	Precision	recall	F1 Score	Support
	0	0.54	1.0	0.70	535
	1	0.93	0.94	0.92	465
	accuracy			0.54	1000
	Macro avg	0.27	0.50	0.35	1000
	Weight avg	0.29	0.54	0.37	1000

Table 11: DeBERTa WITH EPOCH 1

6.4 Experiment / Case Study 4

The library called PYABSA(Aspect based sentiment analysis) has been used for analysing the aspects present in the reviews. Once the dataset is split into the training and the test set. The test set data are fed into the ATEPC interface for finding the aspects and the sentiment present in the review. The comparison between the actual and predicted score for this aspect analysis is not done. But this framework gives the aspects present in the reviews and the sentiment present in it. This model is also built based on the DeBERTa model(Yang and Li, 2022).

6.5 Discussion

Totally four different types of experiments have been conducted in this research. The first three experiments are done using the transformer model like BERT, RoBERTa, and DeBERTa with different epochs. Among these, the DeBERTa model with 2 epochs has achieved an accuracy of 93% and is considered the best model for this dataset. The Second best model is the BERT model and RoBERTa model does not perform well with this dataset. In the fourth experiment, the aspect extraction and the sentiment extraction for these reviews were performed. While manually going through the detected aspects it clearly shows that most of the aspects and sentiments are extracted correctly but also in some of the cases the aspect extraction and sentiment extraction were not correct.

7 Conclusion and Future Work

The main idea of this research is to enhance the detection of fake reviews and aspects using the novel DeBERTa architecture. This research also compares the performance of the other models like BERT, and RoBERTa to the DeBERTa model. The result shows that the DeBERTa model has performed better than BERT and RoBERTa in classifying deceptive reviews. It has achieved an F1 score of 93% and an accuracy of 93 % in two epochs. Aspect and sentiment extraction is also done using DeBERTa. In the future, the same model can be used to find deceptive reviews in larger datasets, and also if any other new pre-trained model is introduced, they also are used along with the already existing model.

References

Abdul, S., Qiang, Y., Basit, S., Ahmad, W., 2019. Using BERT for Checking the Polarity of Movie Reviews. Int J Comput Appl 177. https://doi.org/10.5120/ijca2019919675

Alsubari, S.N., Deshmukh, S.N., Al-Adhaileh, M.H., Alsaade, F.W., Aldhyani, T.H.H., 2021. Development of Integrated Neural Network Model for Identification of Fake Reviews in E-Commerce Using Multidomain Datasets. Appl Bionics Biomech 2021. https://doi.org/10.1155/2021/5522574

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding, in: NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference.

Elmurngi, E., Gherbi, A., 2017. An empirical study on detecting fake reviews using machine learning techniques, in: 7th International Conference on Innovative

Computing Technology, INTECH 2017. https://doi.org/10.1109/INTECH.2017.8102442

Geetha, M.P., Karthika Renuka, D., 2021. Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. International Journal of Intelligent Networks 2. https://doi.org/10.1016/j.ijin.2021.06.005

Gupta, P., Gandhi, S., Chakravarthi, B.R., 2021. Leveraging Transfer learning techniques- BERT, RoBERTa, ALBERT and DistilBERT for Fake Review Detection, in: ACM International Conference Proceeding Series. https://doi.org/10.1145/3503162.3503169

Hai, Z., Zhao, P., Cheng, P., Yang, P., Li, X.L., Li, G., 2016. Deceptive review spam detection via exploiting task relatedness and unlabeled data, in: EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings. https://doi.org/10.18653/v1/d16-1187

Hassan, R., Islam, M.R., 2020. A Supervised Machine Learning Approach to Detect Fake Online Reviews, in: ICCIT 2020 - 23rd International Conference on Computer and Information Technology, Proceedings. https://doi.org/10.1109/ICCIT51783.2020.9392727

He, P., Liu, X., Gao, J., Chen, W., 2020. DeBERTa: Decoding-enhanced BERT with Disentangled Attention.

Hoang, M., Alija Bihorac, O., Rouces, J., n.d. Aspect-Based Sentiment Analysis Using BERT.

Hu, Y., Ding, J., Dou, Z., Chang, H., 2022. Short-Text Classification Detector: A Bert-Based Mental Approach. Comput Intell Neurosci 2022. https://doi.org/10.1155/2022/8660828

Jacob, M.S., Selvi Rajendran, P., 2022. Fuzzy artificial bee colony-based CNN-LSTM and semantic feature for fake product review classification. Concurr Comput 34. https://doi.org/10.1002/cpe.6539

Jindal, N., Liu, B., 2008. Opinion spam and analysis, in: WSDM'08 - Proceedings of the 2008 International Conference on Web Search and Data Mining. https://doi.org/10.1145/1341531.1341560

Karande, H., Walambe, R., Benjamin, V., Kotecha, K., Raghu, T.S., 2021. Stance detection with BERT embeddings for credibility analysis of information on social media. PeerJ Comput Sci 7. https://doi.org/10.7717/peerj-cs.467

Khan, H., Asghar, M.U., Asghar, M.Z., Srivastava, G., Maddikunta, P.K.R., Gadekallu, T.R., 2021. Fake Review Classification Using Supervised Machine Learning, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-68799-1_19

Kim, J., Kang, J., Shin, S., Myaeng, S.H., 2021. Can You Distinguish Truthful from Fake Reviews? User Analysis and Assistance Tool for Fake Review Detection, in: Bridging Human-Computer Interaction and Natural Language Processing, HCINLP 2021 - Proceedings of the 1st Workshop.

Lee, M., Song, Y.H., Li, L., Lee, K.Y., Yang, S.B., 2022. Detecting fake reviews with supervised machine learning algorithms. Service Industries Journal 42. https://doi.org/10.1080/02642069.2022.2054996

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach.

Liu, Y., Wang, L., Shi, T., Li, J., 2022. Detection of spam reviews through a hierarchical attention architecture with N-gram CNN and Bi-LSTM. Inf Syst 103. https://doi.org/10.1016/j.is.2021.101865

Mowlaei, M.E., Saniee Abadeh, M., Keshavarz, H., 2020. Aspect-based sentiment analysis using adaptive aspect-based lexicons. Expert Syst Appl 148. https://doi.org/10.1016/j.eswa.2020.113234

Nasir, J.A., Khan, O.S., Varlamis, I., 2021. Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights 1. https://doi.org/10.1016/j.jjimei.2020.100007

Qasim, R., Bangyal, W.H., Alqarni, M.A., Ali Almazroi, A., 2022. A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification. J Healthc Eng 2022. https://doi.org/10.1155/2022/3498123

Salminen, J., Kandpal, C., Kamel, A.M., Jung, S. gyo, Jansen, B.J., 2022. Creating and detecting fake reviews of online products. Journal of Retailing and Consumer Services 64. https://doi.org/10.1016/j.jretconser.2021.102771

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: Advances in Neural Information Processing Systems.

Yang, H., Li, K., 2022. PyABSA: Open Framework for Aspect-based Sentiment Analysis.