Sentiment Analysis of Customer Reviews Using Deep Learning Techniques

MSc Research Project
MSc Data Analytics

Rishabh Singh Chauhan
Student ID: X21107939

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Rishabh Singh Chauhan		
Student ID:	X21107939		
Programme:	Msc Data Analytics		**Year:2022**
Module:	Research Project		
Supervisor:	Vladimir Milosavljevic		
Submission Due Date:	15 December 2022		
Project Title:	Sentiment Analysis of Customer Reviews Using Deep Learning Techniques		
Word Count:	**5378**	**Page Count:18**	

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: Rishabh Singh Chauhan

Date: 15 December 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):	

Sentiment Analysis of Customer Reviews Using Deep Learning Techniques

Rishabh Singh Chauhan
Student Id: x21107939
Research Project
MSCDAD JAN22 B

December 15, 2022

Abstract

The rapid increase in online shopping leads to the generation of more data like reviews, ratings, and suggestions. This data is very important for businesses to understand their customer's sentiments and how their products are being perceived. The existing machine learning techniques and natural language processing techniques are not able to understand the context of sentences as a whole, which prevents them from providing accurate results for complex problems. Deep learning techniques are therefore required in order to classify and understand reviews more effectively and efficiently. This research has implemented four deep-learning models to analyze and evaluate the best model for sentiment analysis for customer reviews. This research has used Long Short-Term Memory, Bidirectional Recursive Neural Networks, Bidirectional Encoder Representations from Transformers, and Convolutional 1D networks. Performance metrics like Precision, Accuracy, Recall and F1 score have been used for evaluation.

1 Introduction

Online shopping has been significantly increased over the past ten years as a result of the internet's explosive growth and the simplicity with which multiple items can be checked out and bought at once.Today, whether people want to buy a refrigerator or look up restaurant reviews, they can find anything online and shop there with just a click. Every time a customer purchases a product, they have the option of writing a review of it so that other customers can read it and learn more about the product.

Due to the ability to learn about the needs of the customer and how the customer feels about the product, these reviews are helpful not only to the customer but also to the company that makes the product. Through these reviews, businesses were able to learn what customers thought of their products, as well as about the needs of the customer, product flaws, and potential changes that could be made to improve the product and make it more marketable.

1.1 Background

According to the nature of the review, online shopping platforms typically divide customer reviews into three categories: positive, neutral, and negative. It is impossible to manually sort the reviews on the online platform because there are millions of products and billions of reviews that are added every day so it becomes a challenge and an automated system is needed to overcome this problem. Therefore, there is an need of architecture that automatically classifies these reviews according to the nature of the text into different categories i.e., positive, negative, and neutral.In the past, many techniques have been used for the sentiment analysis of reviews starting with the Text blob model, Word-dictionary model, named entity-based sentiment analyzer, and machine learning models but these models have their own pros and cons and they are limited to a particular domain and problem.

1.2 Motivation

Previous research and experiments have demonstrated that machine learning models used by researchers to categorise the sentiments of the reviews perform well for the balanced dataset but fall short for the imbalanced datasets, opening the door for the use of deep learning techniques.This research has proposed deep-learning techniques to classify the sentiments of product reviews using Long Short Term Memory, Bidirectional Recurrent Neural Networks, Convolutional 1D Neural Networks, and Bidirectional Encoder Representations from Transformer.

In a convolutional neural network, each input that is processed has no connection to any other input, so the nodes are not connected to one another. However, In a recurrent neural network with LSTM, each input is connected to every other input through long short-term memory, and as the model is processed, it contains information from the previous sentence that, if necessary, is used in the next sentence.To overcome the dependency problem in Recurrent neural networks due to vanishing gradient LSTM are used as they are feed-backward neural networks. LSTM is used in this research because they process the entire sentence at once rather than breaking it down into individual points and then processing them.
In this research paper Bidirectional RNN and the 1D Convolutional neural network types of architecture are also implemented for the classification of customer reviews. LSTM and GRU are the two types of layers that make up bidirectional RNNs. This allows the model to analyze the more complicated words and improves performance. 1D Convolutional neural network performs the same input transformation at every patch. Any pattern that the model learns can later be recognized at a specific stage, improving model performance. A bidirectionally trained NLP model called a bidirectional encoder representation transformer (BERT) is proposed because it has a greater understanding of a language than a single-trained model does. The BERT model's key characteristic is that it learns the context of each word and every element around it from the beginning to the end of the sentence.

1.3 Research Question and Objectives

Research Question:- How deep learning model can be employed in extracting sentiments from customer reviews to help businesses to gain insight regarding how well the product is being perceived by customers, identify the major issues, and promote their products?
Research objectives:-
1. Examine all the state-of-the-art Sentiment analysis deep learning model
2. Implement Long Short Term Memory, Bidirectional Recurrent Neural Networks, Convolutional 1D Neural Networks, and Bidirectional Encoder Representations from Transformer
3. Evaluate and analyze all the above-mentioned models
4. Discover the model that works best for sentiment analysis of customer reviews
This Research paper is divided into various sections such as Literature Review, Research Methodology and Implementation, Evaluation and Result, Conclusion, and future work for better understanding and ease of reading

2 Literature Review

In the past years, different researchers have used different techniques to classify sentiments. Diverse studies in the field of sentiment analysis have been conducted as a result of improvements in machine learning and deep learning techniques. The various initiatives and works made by various researchers to categorize the sentiments are described in this section. The approaches used for sentiment analysis have been divided up into different sections of the literature, including Machine learning, Convolutional Neural Networks, Deep Neural Networks, Recurrent Neural Networks, Deep Belief Networks, Hybrid Neural Networks, and Recursive Neural Networks.

2.1 Convolutional Neural Network

In this research, the author (Bagheri et al. 2013) proposed a CNN model for the sentiment analysis of the movie reviews dataset. The architecture is a combination of Word2vec and CNN. Word2vec is used in this paper because it has a faster speed even for the larger dataset and its accuracy is high in interpreting the meaning of the word. Three pairs of the convolutional layer along with the pooling layer are used in the architecture. The combination of the seven convolutional layers and word2vec, which is novel in this study, excites readers about the findings. To enhance the model's fit, the Prelu layer is utilized. In order to reduce the co-adaptation of neurons in the nets, dropout regularisation is set to 0.5. The model's accuracy is 45.5%, which is low but acceptable compared to recursive neural networks and matrix-recursive neural networks. The model's low accuracy may be caused by the output's classification into five classes, which could be increased by reducing the output classes.

The author(Rani & Kumar 2019) proposed a CNN model for the sentiment analysis of Hindi movie reviews in this research. A different set of parameters are used to find the best optimum results but as in the result, it can be observed that parameters such as dropout, regularizer, and the number of epochs have not shown any improvements. Only the change in the number of convolutional layers and filter size shows the improvement in the accuracy of the model. The Grid search technique, which provides the best parameters, can be used to accomplish what the author did manually by using a different set of combinations. It has been noted that training time increases along with the number of layers and the size of the filters in the model. The results show that the model with two convolutional layers and filter sizes 3 and 4 outperforms the standard machine learning models, giving an accuracy of 95%.

A combined CNN+LSTM model(Wang et al. n.d.) is suggested in this paper to perform the dimensional sentiment analysis. Unlike earlier CNN models, which treated the entire text as a single sentence, this model divides the text into different regions, with the useful information from each region being weighted in accordance with its significance for the prediction of sentiment. This method makes the model more effective at correctly predicting attitudes and displays cutting-edge findings. The model is applied to the Stanford sentiment treebank dataset and performance metrics such as RMSE, MAE, and Pearson correlation are used to analyze the model. The performance of the model could be boosted by using the parsing technique for structural information identification.

2.2 Deep Neural Network

The author(Baecchi et al. 2016) proposed a novel technique CBOW-DA-LR method which is an extension of the CBOW-LR model for sentiment analysis of short social media texts. By simultaneously learning the vector representation and categorizing the polarity of sentiments in the short text, this model takes into account both the textual and visual representation of the text. When trained with a small training set, the model's performance is initially unsatisfactory, but as the size of the training set increases over time, the model's performance gradually improves. Four publicly accessible datasets are used to apply the provided model, and the results are compared with state-of-the-art models like ESLAM and FSLM in terms of accuracy and F1 measure. The results are encouraging and had a high classification accuracy, outperforming the most advanced model.
This study(Severyn & Moschitti 2015) proposes a convolutional neural network that initializes the network's parameter weights. By doing so, the model is able to optimize only the features that are important to the model and avoid injecting any extra features. The model operates in three steps: (i) filter words with a frequency lower than 5; (ii) refine embeddings using a distant supervision approach, and (iii) initialize the current network using the previously trained parameters. The cost function with l2 norm regularisation parameters is set to solve the overfitting issue. The novelty of this approach is that it combines unsupervised learning of text with supervised data that results in avoiding injecting any extra features in the model.

2.3 Deep Belief Network

The author(Chen & Hendry 2019) proposed a deep belief network to analyze the user rating of the customers from comments. The feature vector is initially created in the input nodes, after which the data is cleaned by removing brief comments. It is then converted into an opinion 50d feature vector and sent to the input layer of the deep belief network. Grid search is used to tune hyper-parameters because it produces the best sets of parameters and the best outcomes. On a different dataset, the model performance varies, but they outperform the baseline models.

2.4 Unsupervised Learning

In this research, the author(*2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing.* 2015) proposed an unsupervised and domain-independent model that detects explicit and implicit reviews. This study offers a model that was trained without using labeled training data. Opinionated words are targeted through multi-word aspects and heuristic rules to detect the aspect type. Pruning methods are applied in such a way that they remove the incorrect aspect. The author compared its results with state-of-the-art results like Wei et al and ARM and founds that the proposed techniques are outperforming the benched techniques. This method uses a relationship between the words in a sentence and the opinionated word to distinguish between explicit and implicit reviews, which is its main attractive feature. This approach could be further improved if clustering methods are used with the proposed architecture.

2.5 Recurrent Neural Network

In this study's analysis, recurrent neural networks and layer-wise relevance propagation are combined to create a model that will produce better results both quantitatively and qualitatively.(Arras et al. 2017) A gradient-based approach is used to apply the LRP method to a bidirectional LSTM in order to determine which word is more responsible for the overall sentiment of the sentence. Important textual patterns in the dataset are found using this architecture, which exhibits the external classifier to train and take the textual explanations through the original classifier.
The authorCan et al. (2018) built a model to overcome the challenge of application of sentiment analysis only to a particular language. The model is applied to reviews that have been translated into English from other languages after it has first been trained using a large dataset of customer reviews and a smaller dataset specific to a given domain. A benchmark of 60% accuracy is used to evaluate this result, and it is also compared to the lexicon-based method. In all four languages, the model can outperform the baseline (60%) performance. Because some words from the languages did not translate into English and remained in the text unchanged, the model was unable to achieve higher accuracy. These words could be dropped to improve accuracy. The author(Sachin et al. 2020) used Recurrent Neural networks and used the baseline models LSTM, and Bi-LSTM. Reviews are first

tokenized and then mapped to a sequence after this data is labeled into encodings of dimension 3. The glove is used for the word embedding matrix. Data overfitting is prevented by the dropout layer with an output size of 128. When LSTM, GRU, and Bi-LSTM were compared to one another, GRU outperformed them all with an accuracy of 74%.By incorporating sentiment logic that is tree-based, the model will perform better because sentiments are more thoroughly analyzed.

2.6 Recursive Neural Network

In order to determine which model is best for the sentiment analysis of Twitter tweets, the author(Yuan & Zhou n.d.) proposed a recursive neural network model with variation in different layers. A model that has been glove-pre-trained is used for text vectorization. The lack of phrases and the dataset's imbalanced effects on the model's behavior makes the experiment's findings unimpressive. By enhancing regularisation and pre-processing the dataset to create a slight balance, model performance may be improved.
This paper(Timmaraju & Khanna n.d.) uses transfer learning to tackle the sentiment analysis issue. The model is first trained on the Stanford sentiment dataset using a recursive neural network, and then it is used as the basis for a recurrent neural network on a different dataset. The parameter cannot be tuned because there are no fine-grained labels. The model's accuracy of 80.8% is quite encouraging for the transfer learning experiment.

3 Research Methodology and Implementation

3.1 Data Collection

The research follows the knowledge discovery in databases methodology. The research methodology is divided into four layers i.e. data collection, data Cleaning layer, data Preparation, and modeling layer. This research has used the Google play store reviews dataset which is available at Kaggle [1]. The dataset consists of 12000 rows and 12 columns which have reviews, and ratings of different apps of the play store. Users have rated and reviewed the Google Play application in the dataset. The dataset is first downloaded in csv format, after which it is mounted on Google Drive so it can be used in Google Collaboratory Notebook for further operations.

[1] https://www.kaggle.com/datasets/therealsampat/google-play-apps-reviews

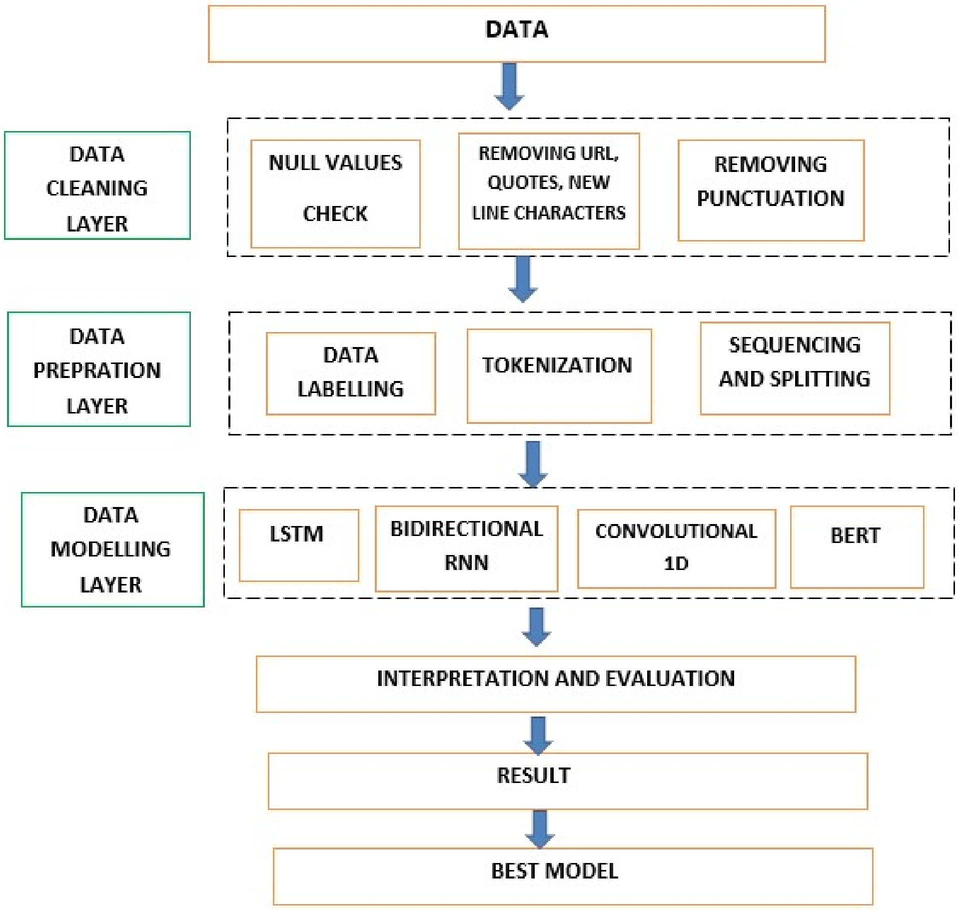

Figure 1: Architecture

3.2 Data Cleaning

3.2.1 Null Values Check

The first thing that is implemented in data cleaning is the checking of null values in the dataset. For this, the .isnull() function is used. In order to maintain the dataset's balance, we removed all the null values.

3.2.2 Removing characters and signs

After checking null values all the URLs, emails, and new lines characters, distracting single quotes from the text are removed and all the text is lowercased.

3.2.3 Removing Punctuations

Punctuations make it difficult for the model to read and analyze the data. Therefore, punctuations are removed from the text using the def sent-to-words(sentences) function and genism library is used to pre-process the text.

	content	new_content
0	I cannot open the app anymore	i cannot open the app anymore
1	I have been begging for a refund from this app for over a month and nobody is replying me	i have been begging for a refund from this app for over a month and nobody is replying me
2	Very costly for the premium version (approx Indian Rupees 910 per year). Better to download the premium version of this app from apkmos website and use it. Microsoft to do list app is far more better.	very costly for the premium version approx indian rupees per year better to download the premium version of this app from apkmos website and use it microsoft to do list app is far more better
3	Used to keep me organized, but all the 2020 UPDATES have made a mess of things !!! Y cudn't u leave well enuf alone ??? Guess ur techies feel the need to keep making changes to justify continuing to collect their salary !!! 😝😝😝	used to keep me organized but all the updates have made a mess of things y cudn t u leave well enuf alone guess ur techies feel the need to keep making changes to justify continuing to collect their salary
4	Dan Birthday Oct 28	dan birthday oct

Figure 2: Data before and after cleaning

3.3 Data Preparation

There are 12 columns in the dataset but only 'content' and 'score' features are needed for this research therefore these two features are extracted into the data frames from the dataset. The content feature has reviews on the applications of the play store and the score feature tells the rating of that review.

3.3.1 Data labeling

The target variable 'score' have values from 1 to 5 therefore the score 1,2 is labelled as "Negative",3 as "Neutral" and 4,5 as "Positive".For this,to-categorical method from Keras library is implemented. Again after converting the score into 'Positive', 'Neutral', and 'Negative' 0 was labelled as 'Negative', 1 as 'Neutral', and 2 as 'Positive'.

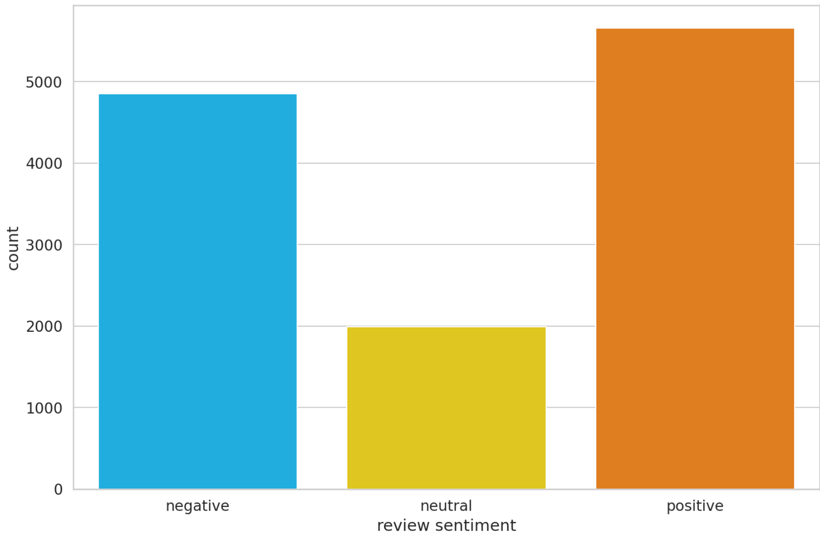

Figure 3: Distribution of data

3.3.2 Tokenization

The process of breaking down sentences into tokens allows for the identification of keywords in each sentence as well as a clear understanding of the meaning of the entire sentence.The sentences are tokenize using the tokenizer function in Keras library

3.3.3 Sequencing and Splitting

To sequence the tokens into the same length padding is done. This improves the performance of the model. After the tokenization and padding part, the data is ready to enter into the model. This data is divided into two parts test and train. In this research the data is split into 75% for training and 25% as a test data. On training, the data model is trained and tuned with parameters, and then the trained model is tested on the test part of the dataset.

3.4 Modelling

After the data has been gathered, prepared, cleaned, and divided, the modeling component of the methodology is applied. In this research, four deep learning models have been implemented and evaluated i.e. Long Short Term Memory, Bidirectional Recurrent Neural Networks, Convolutional 1D Neural Networks, and Bidirectional Encoder Representations from Transformer.

3.4.1 Model 1- Long Short Term Memory

LSTM stands for Long short-term memory. In a Recurrent neural network, the first processing step doesn't affect the output and cost function either therefore LSTM is used in longer sequences where RNN networks face issues due to vanishing gradient. In the LSTM network, the hidden layer has the structure of long and short memory. At any instance t, LSTM gets the feedback on what happened at t-1 instance using the long and short-term memory loops. During the training, the LSTM neuron in the hidden layer learns the specific patterns and associations in the training dataset to make predictions and classifications. The LSTM model is trained on tons of textual data where few words are masked and the model predicts those masked words and keeps learning how to do it smartly like learning associations among words that are close and far apart.

3.4.2 Model 2- Bidirectional Recurrent Neural Networks

This model is the extension of the simple LSTM model. The major difference is that it works bidirectionally which is the reason it performs well on complex sequence classification problems. Bidirectional trains two LSTM units instead of one on an input sequence. It works by connecting the two hidden layers of opposite directions on the same output. The input in the sequence gets information about the past sequence and the future sequence simultaneously which results in better performance of the model. At each time stamp both the encoded representation from backward and forward LSTM is concatenated in the model. This enables the model to analyze more complex patterns than a single LSTM.

3.4.3 Model 3- Convolutional 1d

Convolutional networks can operate in a convolutional manner, which enhances their performance in sequence processing by enabling them to extract features from local input patches. A convolutional 2d neural network is used for image processing and is made up of two layers and every layer has filters of different sizes like 2X2, 3X3, and 5X5, and these filters convolute in two different directions but in convolutional 1d these filters convolute in one direction because the input is one dimension array. Using padding, all the text was first transformed into fixed-length vectors so that each sentence would have the same length in this model. After that, a one-dimensional block that resembles a pixel in an image is placed across the text, and then the results of convolutions of each block of weights on each of the input channels is added to create a single output channel. In this way, Convolutional 1d can have multiple inputs and multiple outputs, and a different set of weights for each channel. As a result, the words will be combined into smaller values, providing information about word pairs and word triplets.

3.4.4 Model 4- Bidirectional Encoder Representations from Transformer

This model is based on the transformer architecture. BERT generates the contextualized embedding meaning of the word. It generates the vector for the whole sentence. The maximum dimension of the BERT vector is 768. The BERT model comes in two different versions, small and large. BERT large has 24 encoders, whereas BERT small has 12 encoders. BERT small has been used in this research because it is compatible with the system and has a shorter run time. BERT is trained on 2500 million words from Wikipedia and 800 million words from books. BERT also offers the pre-process model, which pre-processes the text before it is placed into the BERT layer, for each pre-trained model.

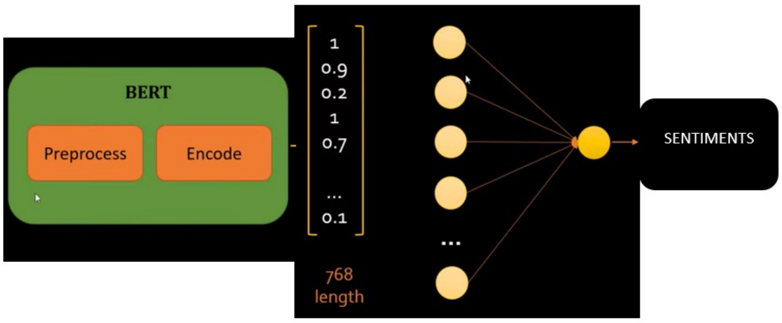

Figure 4: Bert Architecture for Sentiment Analysis

3.5 Implementation

3.5.1 Model 1- Long Short Term Memory

Fifteen hidden layers have been initialsed in this model because the maximum length of the sentence is seven to eight words therefore the number of the hidden layer are doubled according to the number of words in the text sequence. Dropout is used for regularization at a rate of 0.5, closing half the hidden units and deleting each node with a 50% probability during each forward propagation so that the network does not solely rely on neurons but instead learns for itself. The dataset contains a variety of sentiments, including positive, negative, and neutral ones, Therefore SoftMax activation is used due to the multiclass classification of sentiments. For optimization RMS prop optimizer has been used to increase the learning rate and computation. Categorical cross-entropy is used as a loss function because there are three classes and labels are needed to be one hot encoded.

3.5.2 Model 2- Bidirectional Recurrent Neural Networks

For bidirectional RNN same set of parameters are used as implemented in the single LSTM model because both the models have the same architecture and the only difference is the additional LSTM unit in Bidirectional RNN. Additionally, both models use the same data, so there is no change in the parameters.

3.5.3 Model 3- Convolutional 1d

In this model, the vocab size is taken as maximum words, and the dimension of the embedding layer is taken as 40 because if the dimension of the embedding layer is large then there are chances of overfitting in the model, and if the dimensional size is small then there are chances of information loss and the model may underfit. In the convo 1d layer, the kernel size is taken as 6 and the filter size as 20. Five blocks at a time were taken in the max pooling layer, allowing the model to find longer-range dependencies in the convolutions. Global max pooling layers have been utilized to reduce the input representation size. In this model the functional representation is used,so the output from the global max pooling layer is fed to a dense layer, which then feeds the output to another layer using softmax activation. Due to multi-class classification, categorical cross entropy has been used as a loss function and softmax activation is used as an activation function.

3.5.4 Model 4- Bidirectional Encoder Representations from Transformer

Models could be implemented in two different ways: sequentially and functionally. Here,fuctional way is ysed in this model . Firstly an input layer is created in the model with the data type "tf. string," and then this layer is passed through the BERT pre-processing model to get the pre-processed text that can be supply to the encoder. After building the model, the BERT model is fine-tuned by adding a layer of a neural network. The dropout rate was initialised to 0.1. The pooled output has BERT encoding of a 768-size vector because functional style model is created as opposed to a sequential one. Activation function in the second layer is initialized the output of the first layer is placed in the second output layer. After that, the model is compiled using the Adam optimizer and cross-entropy loss function as there aremulticlass in target variable.

4 Evaluation and Discussion

To evaluate the performance of the model performance metrics like Precision, Recall, F1 score, and accuracy have been used.
Precision- is the ratio of true positive or negative predicted by the model to the total number of positive or negative predicted by the model.
Recall- is the ratio of true positives or negatives predicted by the model to the total number of positives or negatives in the data.
F1 score- The harmonic mean of precision and recall is the F1 score. It tells us about the overall performance of the model
Accuracy-The correct positive or negative predicted by the model to the total number of positive or negative outcomes in the data.

Model	Precision	Recall	F1 score	Accuracy
Long Short-Term Memory Model	Positive- 0.77	Positive- 0.81	Positive- 0.79	0.70
	Negative- 0.70	Negative- 0.79	Negative- 0.74	
	Neutral- 0.30	Neutral- 0.16	Neutral- 0.21	
Bidirectional Recurrent Neural Networks	Positive- 0.84	Positive- 0.77	Positive- 0.81	0.71
	Negative- 0.65	Negative- 0.89	Negative- 0.75	
	Neutral- 0.37	Neutral- 0.14	Neutral- 0.20	
Convolutional 1D Neural Networks	Positive- 0.79	Positive- 0.84	Positive- 0.81	0.72
	Negative- 0.66	Negative- 0.89	Negative- 0.75	
	Neutral- 1.00	Neutral- 0.00	Neutral- 0.00	
Bidirectional Encoder Representations from Transformer	Positive- 0.84	Positive- 0.81	Positive- 0.83	0.73
	Negative- 0.75	Negative - 0.79	Negative- 0.77	
	Neutral- 0.38	Neutral- 0.38	Neutral- 0.38	

Figure 5: Result of all the implemented models in the research

Figure 5 tells us about the performance of all the models. Bidirectional Encoder Representations from the Transformer model outperform all the other models with an accuracy of 73% and an F1 score of 83%. The BERT model performs well in predicting neutral reviews with respect to other models. BERT uses transformers that outperform the LSTM architecture. There is also a Bert pre-trained preprocessing model which boosts the model performance. In the results table it is observed that Convolutional 1d neural networks perform well with an accuracy of 73% but fail to predict the neutral reviews with an 0 F1 score. Bidirectional Recurrent neural networks are more effective than Long Short-Term Memory Models

(LSTMs) because they train two LSTMs simultaneously—one in forward propagation and the other in backward propagation.
The order of models according to their performance is given below:-

S.no.	Model Name
1.	Bidirectional Encoder Representations from Transformer
2.	Bidirectional Recurrent Neural Networks
3.	Long Short-Term Memory Model
4.	Convolutional 1D Neural Networks

Figure 6: Order of models according to their performance

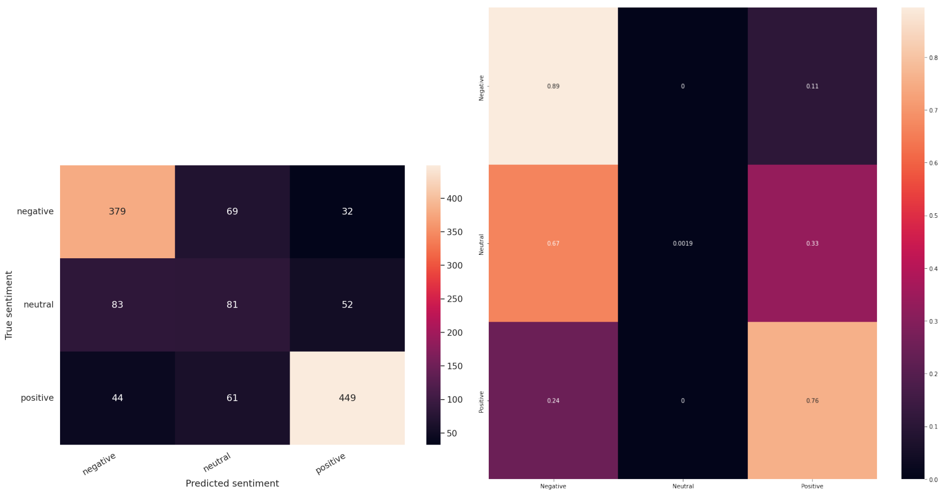

Figure 7:
BERT Heatmap vs. Convolutional 1d Heatmap

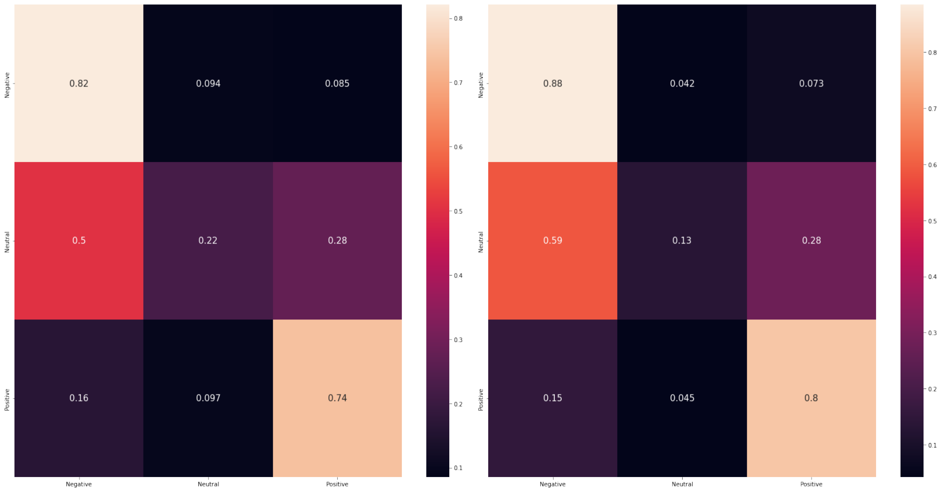

Figure 8:
LSTM Heatmap vs. Bidirectional RNN Heatmap

4.1 Discussion

In the dataset, it can be observed that neutral reviews are less than positive and negative reviews. Therefore in the result table, it can be seen that performance metrics for neutral reviews are low as the model is less trained for a neutral review. Additionally, this research examine how some words in neutral reviews opinionate the model in favor of positive reviews while other words opinionate the model in favor of negative reviews, confusing the model and causing it to predict positive or negative rather than neutral outcomes. In product reviews, the product is either good or bad if the product is good then the majority of reviews in the dataset are positive and if the product is not good then the majority of reviews are negative this creates an imbalance in the dataset and it is very difficult for the model to predict the correct sentiment of the reviews. To overcome this imbalance dataset problem either the data is undersample or oversample to make the balance between different features of target class. This imbalance dataset problem is resolved by the BERT model as they use transformers in their architecture to predict the outcomes. In this research, it is also analyzed that there are type 1 and type 2 errors in the confusion matrix of the predicting models which are decreasing the performance of the model.

In this research among all the models, Bidirectional Encoder Representations from Transformer perform the best therefore they can be used by businesses to predict the sentiments of the customers. The BERT model is currently used by google to optimize the interpretation of user search queries but this model can also be deployed by large businesses in their ecosystem to analyze and predict the sentiments of the customer. With the use of BERT architecture and technology, businesses can gain insight into how well the product is being perceived by customers, identify the major issues, and promote their products through the reviews given by the customer on their products.

5 Conclusion and Future Scope

This research examines the sentiment analysis of customer reviews through deep learning techniques and how these techniques could be beneficial for businesses to predict the sentiments of their customer and product. Four different deep learning models are implemented and their output is evaluated through different performance metrics. Bidirectional encoder representations from the transformer(BERT) model outperforms all the other model with an accuracy of 73% and an f1 score of 83%. This study founds that the bidirectional recurrent neural networks perform similar to the BERT model for positive reviews, but inadequately for neutral reviews. Due to the low number of neutral reviews in the dataset, all models aside from the BERT model are unable to predict them. The keywords in neutral reviews with mixed opinions cause the model to predict both positive and negative outcomes. The BERT model, which also functions with imbalanced data, can be used in businesses to forecast customer sentiment which will help businesses hto gain insight regarding how well the product is being perceived by customers, identify the major issues, and promote their products.BERT small model which has 12 encoders are used in this research but in the future BERT large model with 24 encoders can be used to predict the sentiments with more accuracy and precision. Generative pre-trained transformer 3 model is the most advanced model for human like text classification but this model is not open source and has not very easy access so in future when this model is open source it can used for sentiment analysis of customer reviews.

6 Aknowledgement

I would like to thank my mentor Vladimir Milosavljevic for guiding me throughout the course. I am grateful to my friends and family for their support.

References

2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. (2015), IEEE.

Arras, L., Montavon, G., Müller, K.-R. & Samek, W. (2017), 'Explaining recurrent neural network predictions in sentiment analysis'.
URL: *http://arxiv.org/abs/1706.07206*

Baecchi, C., Uricchio, T., Bertini, M. & Bimbo, A. D. (2016), 'A multimodal feature learning approach for sentiment analysis of social network multimedia', *Multimedia Tools and Applications* **75**, 2507–2525.

Bagheri, A., Saraee, M. & Jong, F. D. (2013), 'Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews', *Knowledge-Based Systems* **52**, 201–213.

Can, E. F., Ezen-Can, A. & Can, F. (2018), 'Multilingual sentiment analysis: An rnn-based framework for limited data'.
URL: *http://arxiv.org/abs/1806.04511*

Chen, R. C. & Hendry (2019), 'User rating classification via deep belief network learning and sentiment analysis', *IEEE Transactions on Computational Social Systems* **6**, 535–546.

Rani, S. & Kumar, P. (2019), 'Deep learning based sentiment analysis using convolution neural network', *Arabian Journal for Science and Engineering* **44**, 3305–3314.

Sachin, S., Tripathi, A., Mahajan, N., Aggarwal, S. & Nagrath, P. (2020), 'Sentiment analysis using gated recurrent neural networks'.

Severyn, A. & Moschitti, A. (2015), Twitter sentiment analysis with deep convolutional neural networks, Association for Computing Machinery, Inc, pp. 959–962.

Timmaraju, A. & Khanna, V. (n.d.), 'Sentiment analysis on movie reviews using recursive and recurrent neural network architectures'.

Wang, J., Yu, L.-C., Lai, K. R. & Zhang, X. (n.d.), 'Dimensional sentiment analysis using a regional cnn-lstm model'.

Yuan, Y. & Zhou, Y. (n.d.), 'Twitter sentiment analysis with recursive neural networks'.