CrossSenti: A Zero Shot Cross Domain Sentiment Classification of Customer Reviews Using Deep Learning Frameworks

[1]Ajeet Kumar and [2]Kumar Abhishek

[1]Department of Computer Science & Engineering, National Institute of Technology Patna, Bihar, India

Bakhtiyarpur College of Engineering Bakhtiyarpur Bihar India

[1]azit.bce@gmail.com

[2]Department of Computer Science & Engineering, National Institute of Technology Patna, Bihar, India

[2]kumar.abhishek.cse@nitp.ac.in

Abstract:

Cross-domain text sentiment analysis (CDTSA) is the process of identifying the feeling emotion opinion attitude also called sentiment of the user comments or reviews written in text across multiple domains, such as reviews, social media posts, or news stories, when sentiment distribution and expression change dramatically between domains. Since conventional sentiment analysis models are usually trained and tested on data from a single domain. When sentiment analysis models are applied to domains other than those for which they were trained, their performance frequently deteriorates. In this paper, we propose a zero shot cross-domain sentiment analysis using a deep learning framework, where we trained the model on a source domain and evaluated in a distinct target domain. Our proposed method does not need to retrain the model for the target data. We demonstrate the global vector (GloVe) word embeddings method that quantifies the text data and combines with a convolutional neural network (CNN) and Multi-channel CNN-based models. We test the models on 12 pairs in cross-domain of amazon customer review datasets and perform an assessment on metric accuracy, precision, recall, and F1-score. And our proposed method outperforms the state-of-the-art results on cross-domain amazon review data sets. The experimental assessment of the cross-domain dataset indicates that the proposed strategy attains remarkable performance.

Keywords: Cross-domain, CNN, RNN, Multi-channel CNN, Word2Vec, GloVe, FastText, CDTSA

1. INTRODUCTION

Sentiment analysis is the branch of natural language processing (NLP) that examines user-generated material, such as reviews, on websites. These evaluations might be ratings, words, sentences, or documents. When Sentiment analysis applied within a single domain such as product reviews or movie reviews, it works well but faces limitations when trying to generalize across different contexts or industries. Cross-domain sentiment analysis addresses these limitations by enabling models to better understand sentiment in diverse domains or areas of application. Cross-domain text sentiment analysis refers to the task of analyzing sentiment expressed in text data from one domain using a model that was trained on data from a dissimilar, often unrelated, domain. The key challenge in cross-domain sentiment analysis is that different domains (e.g., product reviews, movie reviews, and social media posts) often exhibit distinct vocabulary, expressions, and writing styles, making it difficult for a model to generalize across them.

The present estimated e-commerce value is US$ 13 trillion and is projected to achieve US$ 55.6 trillion by 2027, according to a report from Business Wire. Reviews of products and services are important when making decisions about what to buy in this rapidly growing market. Internet users share their opinions and content on social media and websites for the purpose of exchanging information or expressing themselves about a range of subjects, from politics to goods and services. Popular social media platforms include Instagram, Twitter, and facebook. The public generally views user-generated content as authentic and trustworthy to confirm the efficacy of particular products and services [1], since it is typically created by users based on their own experiences. That means online reviews written by users have more impact than reviews written by experts [2]. This verdict is in line with [3] who said that online reviews and comments influence 90% of consumers' purchasing decisions.

From the aforementioned, it can be inferred that businesses should be aware of the sentiment polarity of these online reviews and comments regarding a product. By using the data, marketers can make more informed decisions that will direct their marketing strategies toward the interests and preferences of their target audience and increase revenue [4]. This necessity has led to the emergence of a new domain in information processing, commonly referred to as sentiment analysis [5]. Liu et al [6] describe sentiment analysis as ‘‘the computational process of recognizing and classifying various opinions (thoughts or judgements) expressed in texts’’ By the definition, goal of sentiment analysis is to assess the polarity of online reviews and comments in order to determine if the public has a positive, negative, or neutral opinion about a product or problem [7].

However, most sentiment analysis research initiatives presently depend on a singular data source, such as Twitter or online review platforms. These datasets may exhibit bias and may not truly represent the views of the broader populace regarding the subject, product, or service [4], [7], [8]. As a result, sentiment analysis has recently progressed, utilizing multiple data sources instead of just one. The number of labelled datasets required to train a sentiment classifier may rise when several data sources are used for a given area of interest. Sentiment analysis frequently encounters the issue of an inadequate number of labelled datasets. In this paper we use multi-domain sentiment analysis also referred to as cross-domain sentiment analysis, is used to address the issue.In cross-domain sentiment analysis, sentiment knowledge will be transferred from a resource-rich domain to a resource-limited domain. Because opinions are expressed differently in different domains [2], [9], [15], [18], and [19], sentiment analysis is a domain-dependent process [17]. For example, the term "easy" has a negative connotation in the Books domain but a favorable connotation in the Electronics domain. Thus, creating enough labeled datasets for each domain to train a classifier is the simplest method to solve this issue [2].The drawback of this approach is the high expense of manually labeling enough samples across all potential domains of interest [2], [19].Additionally, a domain-dependent classifier needs a lot of labeled datasets to be accurate[7], [15], [17], [22].

An improved approach to address this issue involves utilizing resource-abundant characterized data from several domains and developing a additional resilient sentiment classifier capable of functioning across various domains [15], [17], [19], [22].The situation in which a classifier operates across various domains is referred to as multi-domain or cross-domain sentiment analysis [19], [21], [22]. Cross-domain sentiment analysis is advantageous when acquiring labelled information proves challenging.

As aforementioned, the comprehension and examination of reviews play a crucial role in the process of making informed purchasing choices. Both negative and positive judgments hold equal significance. According to a research survey, a significant majority of customers, specifically 82%, actively seek out bad evaluations while making purchasing decisions. The internet marketplace, with a robust economy of 13 trillion, is heavily influenced by the peer effect, making reviews a pivotal factor in consumer decision-making over purchases. By utilizing Natural Language Processing (NLP), individuals are able to automate the task of analysing reviews. This study investigates a range of Machine Learning and deep learning techniques in order to forecast the cross domain sentiment of reviews on e-commerce websites. The primary contributions of this work are:

- Compilation of publicly accessible assessments of raw datasets which include both Amazon product reviews and associated metadata.
- Pre-processing of raw dataset.
- Apply different word embedding methods like Word2Vec, GloVe, and FastText to represent dataset into word vector.
- Design a deep learning framework to decide the polarity of different reviews.
- Analyse performance of proposed model on the basis of accuracy precision recall and f1_score in cross domain.

The subsequent sections of the paper are well-organized in the following manner. The section titled "Related work" provides an overview of the background information. The section titled "Methodology" presents the methodology employed in related studies. The section titled "Experimental analysis and Results" presents the findings, followed by the conclusion and future study.

2. RELATED WORK

Sentiment analysis is the examination of a service provider's attitudes about the service they have used, whether they are positive or negative polarity, like or dislike, and love or hate. The sentiment about any product or services is available in word, sentence, document, and ratings provided by the user. These reviews are analyzed by a number of machine learning approaches to analyze the sentiment whether it is positive or negative. Several researchers had worked to classify the sentiments which are described here. Kongthon et al [23] developed online help desk system that utilizes natural language processing and artificial intelligence techniques to offer customer service in a far more interactive and cost-effective manner than with conventional techniques. The presented system demonstrates an e-commerce service platform which increases their customer satisfaction and retention. Jean et al [24] provide a method that relies on consequence sampling to enables the utilization of a large-scale vocabulary within the Neural Machine Translation (NMT) model without escalating the intricacy of the training process. However, there could be a lot of obstacles in the way of creating, optimising, or creating a workable approach for Natural Language Processing (NLP). This approach fails to leverage contextual sentimental knowledge pertaining to the specific feature. Liang et al. [25] introduce a graph convolutional network that utilizes SenticNet to exploit the emotive relationships inside sentences, focusing on a particular component. The author constructs graph neural networks by incorporating SenticNet's emotive information in order to enhance sentence dependency graphs.

In their investigation, Strubell et al. [26] employed substantial quantities of unlabeled data. Previous research has demonstrated that the integration of natural language processing (NLP) with a neural network model has generated favourable accuracy outcomes. Furthermore, the extent of accuracy enhancement is contingent upon the allocation of computational resources. The author has provided cost-cutting tips based on thorough study.

In a similar vein, the utilization of natural language processing (NLP) techniques to examine data from the realms of accounting, auditing, and finance is being employed in order to extract valuable insights and draw inferences that contribute to the generation of knowledge. Fisher et al. [27] have conducted a study employing Natural Language Processing (NLP) techniques within the field of accounting, and have also outlined potential avenues for further research. In addition to the aforementioned contributions, Vinyals et al. [28] have proposed a novel approach to address the challenge of variable-size output dictionaries.

Natural Language Processing (NLP) approaches have been employed in standardized dialogue-based systems, such as chat boxes [29]. Text Analytics is a widely employed domain in which Natural Language Processing (NLP) is extensively utilized [30]. Machine learning methods in conjunction with

natural language processing (NLP) have the potential to be employed for additional purposes such as translation, summarization, and data extraction. However, it is important to note that these applications often come with significant computational expenses.

Deep learning has emerged as a significant tool in the prediction of diseases, including COVID-19 and other ailments, during the ongoing pandemic [31][32][33]. The textbook "Deep Learning for NLP and Speech Recognition" provides a comprehensive exploration of the theoretical side. This article provides an elucidation of the Deep Learning Architecture, specifically focusing on its applications in various Natural Language Processing (NLP) Tasks. Additionally, it establishes a connection between deep learning approaches and NLP as well as speech and text processing. Furthermore, it offers practical guidance on effectively utilizing the tools and libraries associated with deep learning in real-world scenarios.

The identification of polarity in text as part of sentiment analysis is a crucial undertaking within NLP methodologies. To determine polarity, researchers employed unsupervised and reproducible sub-symbolic methods, including auto-regressive language models, which converted spoken language into a protolanguage format [34]. The concept of polarity is a fascinating framework for understanding the nuanced nature of sentiments. Trueman et al. [35] introduced a novel approach to enhance sentiment analysis, in this method employs a convolution-stacked bidirectional long short-term memory model integrated with a multiplicative attention mechanism. This method seeks to accurately determine sentiment polarity and aspect categories. The field of affective computing and sentimental analysis, which encompasses the study of human-computer interaction, machine learning, and multi-model signal processing, has been suggested as a means to comprehend the significance of individuals' sentiments expressed on social media platforms [36]. The sentiment data obtained occasionally exhibits imbalances and insufficiencies. The issue of inadequate and unevenly distributed data is tackled by the meta-based self-training approach, which incorporates a meta-weighted known as MSM (Meta-based Self-Training Method with Meta-Weighted) [37]. The MSM model is founded upon neurosymbolic learning processes. A subsequent investigation was conducted to assess the potential bias exhibited by the pre-trained learning model utilized for sentiment analysis and emotion detection [38]. We have briefly described some related literature in tabular form as shown in table 1.

Table 1 Related reviews summary

References	Used Datasets	Applied Algorithms	Method	Embedding /Features	Accuracy (%)
[39]	Reviews of 15,000 user collected from different websites	Neural Network	Supervised	POS Tagging	80
[40]	positive and negative classes of 300 reviews	KNN, NB, DT	Supervised	-	71
[41]	Positive, negative, and neutral classes of 3241 sentences	NB,LR,SVM,RCNN,Hybrid Multichannel Approach	Supervised	Glove,Word2Vec and FastText	82
[42]	three classes of 10021 user reviews from different websites	Rule based, N-gram and RCNN	Both Supervised and Unsupervised	Word2Vec	75
[43]	11,000 user reviews of positive and negative classes	NB,KNN,ANN,LR,DT,SVM,RF,AdaBoost,wVoting	Supervised	n-gram	82
[44]	11,000 user reviews of Positive and negative classes	NB,KNN,ANN,LR,DT,SVM,RF,AdaBoost,wVoting	Supervised	n-gram	81
[56]	22127 comments on celebrities(in-domain)	LSTM-CNN	Supervised	TF-IDF	76

[56]	1590 comments on talent show (cross-domain)	PersBERT	Supervised	TF-IDF	68

Training and testing datasets must adhere to the same distribution to employ conventional machine learning techniques. All study included in Table 1 has evidenced the effectiveness of the supervisory classification method in sentiment analysis. This represents a distinct categorization for text sentiment analysis, indicating that the text is classed based on the viewpoint of a certain topic. Text sentiment analysis is generally categorized into three classes according to different levels of textual granularity: phrase-level sentiment analysis, sentence-level sentiment analysis, and document-level sentiment analysis. The cross-domain text sentiment analysis difficulty occurs due to the significant decline in performance of supervised classification algorithms when the labelled data does not correspond to the domain of the data being assessed.

The difficulty of knowledge transfer across different domains is an aspect of cross-domain textual sentiment analysis. Meng et al. [46] recommend a machine learning-based cross-domain text sentiment analysis technique from the viewpoints of features and samples. The uses of expectation maximization (EM), support vector classification (SVC), and naïve Bayes (NB) in cross-domain text sentiment analysis are contrasted by Huang et al. [47]. According to the experimental findings, the EM approach outperforms the NB and SVC methods by a small margin. Xia et al. [48] create an ensemble model for the NB, SVC, and EM approaches using ensemble features such as word relation and part-of-speech tagging; the experimental results surpass those of traditional single machine learning techniques. Tang et al.'s [49] prior research indicates that deep learning techniques surpass conventional methods in sentiment categorization, view the extraction procedure, and emotion dictionary development. Yu et al. [50] utilize a deep learning methodology to model words, and the experimental results indicate that the model created using deep learning surpasses the traditional framework associated with learning models. Deep learning often surpasses conventional graph models and statistical learning methods due to its superior ability to reveal and acquire the intrinsic semantic representation of textual content across many domains.

After reviewing all of the aforementioned papers, we observed that several research have been published in which the source and target domains are same. A few numbers of researches is available for cross domain distribution.We have presented a comprehensive framework for sentiment analysis across diverse domains, wherein the system is trained on the source domain and shares the weights of the feature vector from the source domain framework to extract features that are pertinent in the target domain. And also use the variant word embedding model, a productive technique that extracts syntactic and semantic information two crucial aspects of sentiment analysis from vast amounts of unstructured text data by learning high-quality vector representations of words.

3. METHODOLOGY

The comprehensive methodology utilized for sentiment recognition in reviews is depicted in Figure 1, depicting the block diagram. The process of sentiment detection in reviews involves three primary stages: The three main components utilized in this study are data pre-processing, word embedding, and the employed models.

Data Pre-processing:

Pre-processing techniques are an essential step in cross domain text sentiment analysis to convert unprocessed textual data into machine learning model compatible format. Here are some common pre-processing techniques used in cross domain text sentiment analysis:

Tokenization: This entails deconstructing the text into discrete words or tokens. This phase is essential for transforming the unrefined text input into a format suitable for processing by machine learning

algorithms.
Stop words Removal: Common words like "a," "an," "the," "and," and "of" that don't have much sense in the text are known as stop words. Eliminating stop words can help machine learning models perform better by lowering the dimensionality of the data.

Stemming and Lemmatization: Reducing words to their fundamental form is known as stemming. (e.g., "running" to "run"), while Lemmatization is the process of breaking down words into their most basic form. (e.g., "running" to "run" and "ran" to "run"). These methods can assist in lowering the data's dimensionality and enhancing sentiment analysis's precision.

Spell checking: This involves correcting spelling errors in the text, which can decrease data noise and increase sentiment analysis accuracy.

Part-of-speech (POS) tagging: Finding each word's part of speech—noun, verb, adjective, etc.—in a document is known as POS tagging. This information can be used to identify sentiment-bearing words and improve the accuracy of sentiment analysis.

Handling negation: Negation handling involves identifying negation words such as "not" and "never" and their scope in the text. This is important because negation can reverse the polarity of sentiment words and phrases.

Encoding: Finally, the pre-processed text data is typically encoded using word embedding techniques such as Word2Vec or GloVe to represent the text as a set of numerical vectors that can be processed by machine learning models.

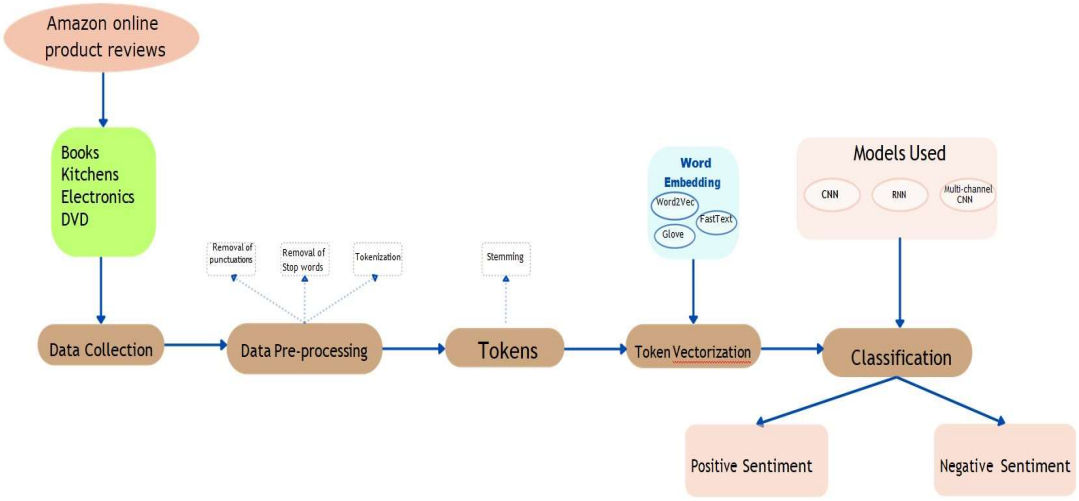

Fig 1: Flow chart of the proposed model.

Word Embedding:

Text classification comprises two fundamental steps: initially, transforming text into numerical representations through a word embedding technique, and subsequently, applying these numerical interpretations to machine learning techniques or deep learning architectures for further analysis. This research employs three methods: word2vec embedding, GloVe vector embedding, and fastText vector embedding. Word embedding is the process of turning text into a numerical representation that is vectorized. Nearly all deep learning architectures and many machine learning algorithms are unable to process plain text or raw strings or perform any kind of task, including classification and regression. Generally speaking, they require numerical inputs. Furthermore, it is essential to extract information from the vast volume of data in text format and to develop applications.

Word2Vec Embedding:

First, Google researcher Tomas Mikolov presented word2vec, a variant of word embedding

techniques, in 2013 [51]. It has vectors of words that are comparable. To put it another way, it finds similarities mathematically. Word2vec generates vectors to disseminate numeric representations of word attributes, such as the word context. This study uses the skip-gram technique due to its superior accuracy in analysing bigger, high-dimensional datasets. The skip-gram method utilizes the target word as input to predict the outputs of the words surrounding it, informed by the context. In the phrase "I have a cute dog," the input is "cute," and the output comprises "I," "have," and "dog." The size of the window is presumed to be 5.

The length of the window governs the word span on both sides of target word that could be considered context words. The primary goal of skip-gram model training is to optimize and improve the likelihood of predicting the appropriate context words for the given target word. The goal could be expressed as the mean of the log probability for a sequence of training words $w_1, w_2, \ldots, w_N$. The log probability (l_p) is computed as follows:

$$l_p = \frac{1}{N}\sum_{n=1}^{N}\sum_{-c\le j\le c, j\neq 0}\log p(w_{n+j}|w_n) \quad (1)$$

Where c denotes the size of the training context window, which varies depending on the centre word w_n. A larger c generates more training examples, which can leads to improved accuracy. The simple skip-gram method computes this probability $p(w_{n+j}|w_n)$ using an activation function called softmax activation as follows:

$$p(w_O|w_I) = \frac{\exp({v'_{w_O}}^{N} v_{w_I})}{\sum_{w=1}^{W}\exp({v'_{w_O}}^{N} v_{w_I})} \quad (2)$$

Where W is the vocabulary dimensions, v is the target that is input and $v'$is context that is output vector representations of the word, respectively.

GloVe Embedding:

Glove [52] is built on a matrix of factorization algorithms on the word-context matrix and was derived from Global Vectors. Additionally, this model is a kind of unsupervised learning technique that generates word vector representations. Initially, it creates a sizable matrix of co-occurrence information (words × context); that is, you count the number of times each "word" (the rows) appears in a huge corpus's "context" (the columns). GloVe has the advantages of not relying solely on local data as word2vecdoes, but also on global data to obtain word vectors.

The conditional probability $p(w_O|w_I)$ between any two words w_O and w_I in the dictionary D based on the GloVe vector model is computed using a softmax function as shown in equation 2. Here the Global Vector model makes the three changes to the skip-gram version, all of which are based on square loss.

- Use non-probability distribution variables $q'_{ij} = \exp({v'_{w_O}}^{N} v_{w_I})$ and $p'_{ij} = x_{ij}$ and take the logarithm of both, so the square loss term is $({v'_{w_O}}^{N} v_{w_I} - \log x_{ij})2$.
- For each word $w_i$add two scalar model parameter: the centre word bias b_i and the context word bias c_i .
- Substitute the weight function $h(x_{ij})$ for each loss term, where $h(x)$ increases in the interval $[0, 1]$.

Finally, when we combine all of the changes, we train GloVe to minimize the following loss function:

$$\sum_{i\epsilon W}\sum_{j\epsilon W} h(x_{ij})\,({v'_{w_O}}^{N} v_{w_I} + b_i + c_i - \log x_{ij})2 \quad (3)$$

These non-zeros x_{ij} represent precomputed global corpus information, which is why the model is called GloVe for Global Vectors.

Fast Text Embedding:

In 2016, Facebook researchers proposed fastText embedding as an extension of word2vec [53]. This model is a technique for unsupervised learning that generates word vector representations. Facebook

developed pre-trained models of fastText that is available for 294 languages. Therefore, FastText goes one step farther than simply feeding the single word into neural network. FastText, on the other hand, divides words into several sub-words (n-grams) in form of parts of words and the characters. The sum of these n-grams will be the word vector embedding for word. Given the training dataset, the neural network will have word embeddings for every n-gram after training. Since it is very likely that some of the n-grams of rare words also occur in other words, rare words can now be represented appropriately.

4. MODELS USED:

After the words were embedded, three distinct deep learning models were used for sentiment detection: a CNN, an RNN and a CNN with multiple input channels.

Convolutional neural network (CNN): a convolutional neural network (CNN) is used for tasks that require processing images. It has been put to use in a variety of natural language processing applications, including spam detection, language modelling, and sentiment categorization. Now, we typically associate neural networks with the products of matrices, although this isn't always the case with CNN. It uses a special method called Convolution. When two functions are convolved, a third function is generated that shows how the first function was altered by the second. A framework of CNN architecture for cross domain sentiment analysis and its pseudo code are below respectively. There are three components that make up a convolutional neural network:

- **Convolution Layer:** This layer is used to extract various features from the input data by using a sliding window or feature matrix with a non-linear activation function.

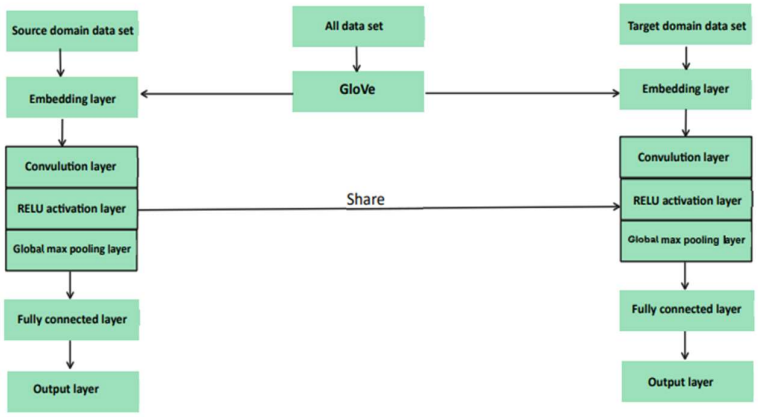

Fig 2: Framework of CNN model.

- **Pooling Layer:** This layer is used to reduce the input dimensions of feature matrix while producing a fixed output matrix in order to reduce computational costs. It accepts input matrices of various sizes. This is achieved by working separately on each feature map and minimizing the connections between layers.
- **Fully Connected Layer:** This layer utilizes data from the pooled output layer to produce the final output, which ascertains whether the review is good or negative. This layer constitute the final segments of the CNN architecture and are often situated prior to the output layer.

Algorithm 1 **Pseudocode for CNN model**

Input: Source Dataset S_D, Target Dataset T_D

Output: Sentimentclass (pos or neg)

1: //Training

2: for each $R \in S_D$ do:

3: Pass R into embedding layer

4: A Convolutional layer receives the output of the preceding layer

5: Then a GlobalMaxPooling is applied

6: Then the result is passed from a dense layer for binary Classification

7: end for

8: Split the T_D into T_{DTune} and T_{DTest}

9://Tuning

10:for each $R \in T_{DTune}$ do

11: Pass R into embedding layer

12: A Convolutional layer receives the output of the preceding layer

13: Then a GlobalMaxPooling is applied

14: Then the result is passed from a dense layer for binary Classification

15: end for

16: //Testing

17: for each $R \in T_{DTest}$ do

18: Classify R into sentiments using trained model.

19: Show Result.

20: end for

Based on sentiment categorization, the CNN model can be described as a sequential model consisting of five layers. The architectural design comprises an input layer with a size equivalent to its length. The subsequent layer, known as the embedding layer, is implemented on top of the initial layer. This layer is composed of a vocabulary size of 100,000, an embedding dimension of 100, an input length of the maximum length (100), and incorporates an embedding matrix as its weights. The subsequent layers comprise a 1-dimensional Convolutional layer positioned above the embedding layer, featuring a filter size 32 and a kernel size 4. The rectified linear unit (ReLU) is used as activation function utilized in this layer. Following the application of the 1D convolutional layer, the subsequent step involves utilizing the global max pool 1D layer for pooling. Following the pooling layer, the further steps involve the utilization of two dense layers. The penultimate layer consists of 256 neurons and the rectified linear unit (ReLU) is used as activation function. The last output layer, on the other hand, comprises a single neuron and employs the sigmoid activation function. The aforementioned model is created with the Adam optimizer, 'binary_crossentropy' loss function, and accuracy measures.

***Recurrent Neural Network (RNN)*:** The utilized model is a three-layer sequential model based on Recurrent Neural Networks (RNNs). The initial layer is the embedding layer, which is comprised of a vocabulary size of 100,000, an embedding dimension of 100, an input length of the maximum length, specifically 100. Additionally, it includes an embedding matrix that serves as the weights for this layer. The second layer is sent through a SimpleRNN layer consisting of 100 neurons. The output generated by the second layer is subsequently sent into a dense layer, which consists of a single neuron and employs the activation function as 'sigmoid'. The aforementioned model is ultimately constructed with the Adam optimizer, 'binary_crossentropy' as loss function, and accuracy metrics.

Subsequently, the Multi-channel Convolutional Neural Network (CNN) was employed, exhibiting notable similarities to the preceding model. Figure 3 depicts a visual representation of a Recurrent Neural Network (RNN).

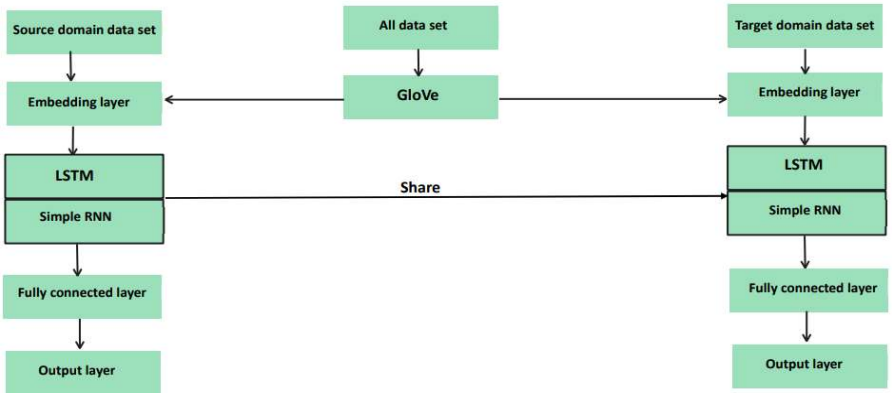

Fig 3: A Framework of RNN Model

Algorithm 2 Pseudo code for RNN model

Input: Source Dataset S_D, Target Dataset T_D

Output: Sentimentclass (pos or neg)

1: //Training

2: for each $R \in S_D$ do:

3: Pass R into embedding layer

4: A LSTM layer receives the output of the preceding layer

5: A SimpleRNN layer receives the output of the preceding LSTM layer

6: then the result of simpleRNN is passed from a dense layer for binary classification

7: end for

8: Split the T_D into T_{DTune}andT_{DTest}

9://Tuning

10:for each $R \in T_{DTune}$ do

11: Pass R into embedding layer

12: A LSTM layer receives the output of the preceding layer

13: A SimpleRNN layer receives the output of the preceding LSTM layer

14: then the result of simpleRNN is passed from a dense layer for binary classification

15: end for

16: // Testing

17: for each R in T_{DTest} do

18: Classify R into sentiments using trained model

19: show results

20: end for

Multi-channel CNN: The study utilizes a model of three channels, each channels display identical

architectural features. Each channel consists of an initial layer, known as input1, which is referred to shape corresponding to its length. The next layer is known as embedding layer, has 100 neurons and a vocabulary size. It is applied to the first layer. A Conv1D layer is subsequently employed, including a filter size 32, a kernel size 4, and utilizing activation function as ReLU. The rate 0.5 of dropout layer is added to the Conv1D layer. Then a max-pooling layer is introduced with a pooling dimension of two. The resultant output is subsequently flattened and placed in a one-dimensional layer. Channels 2 and Channel 3 display an identical layer sequence, utilizing attribute values consistent with those in channel 1. The outcomes of channel 2 and channel 3 are augmented and thereafter archived in layers flat 2 and flat 3 in order. The data from flat 1, flat 2, and flat 3 is eventually amalgamated and the outcomes stored in the merged layer. Subsequent to obtain the result from the amalgamated merged layer, two substantial layers were utilized. The primary dense layer comprises 10 neurons, each employing the activation function as rectified linear unit (ReLU). This is followed by another thick layer containing a single node that utilizes the sigmoid activation algorithm. A model is ultimately formed using input1, input2, and input3, while employing the outputs produced by the final dense layer. The model is compiled with the binary cross-entropy loss operation, the ADAM optimization, and accuracy measures. Figure 4 illustrates the architectural design.

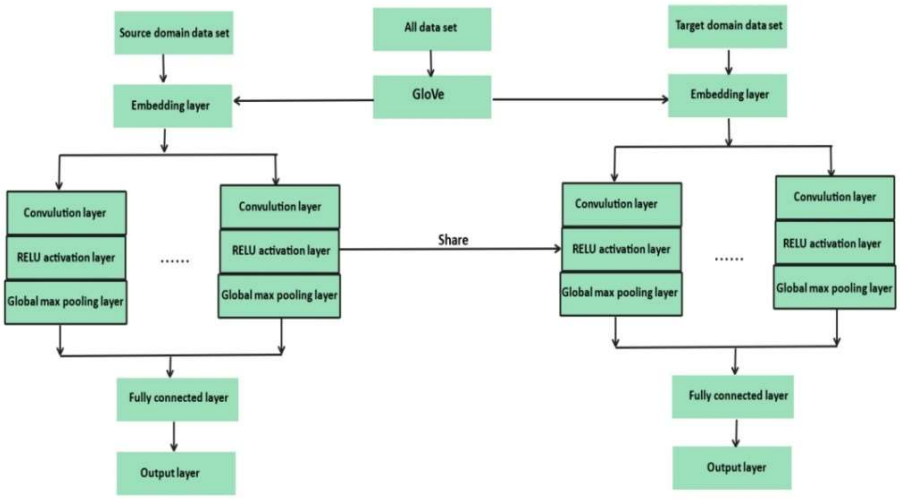

Fig 4: A Framework of Multi-channel CNN Model.

ALGORITHM 3 Pseudo Code For multi-channel CNN MODEL

Input: Source Dataset S_D, Target Dataset T_D

Output: Sentiment class (pos or neg)

1: //Training

2: for each $R \in S_D$ do:

3: Pass R into embedding layer

4: A Convolution layer receives the output of the preceding layer

5: then perform GlobleMaxPool

6: then the ouput of each channel is merged together

7: then the merged output is passed from a dense layer for binary classification

8: end for

9: Split the T_D into T_{DTune} and T_{Dtest}

10: //Tuning

11: for each $R \in T_{Dtune}$ do

12: Pass R into embedding layer

13: A Convolution layer receives the output of the preceding layer

14: then perform GlobleMaxPool

15: then the ouput of each channel is merged together

16: then the merged output is passed from a dense layer for binary classification

17: end for

18: //Testing

19: for each R in T_{Dtest} do

20: Classify R into sentiments using trained model

21: show results

22: end for

5. EXPERIMENTAL ANALYSIS AND RESULTS

We have gone through rigorous experimentation and presenting the corresponding outcomes.This section provides a comprehensive overview and critical analysis of the dataset description, experimental setup, and experiment outcomes. We conducted extensive test on electronics kitchens DVD and books datasets to test our proposed model performance for sentiment analysis of a review. We have trained the model on one dataset and test the model on remaining three datasets. The precision, recall, F-measure, and accuracy metrics were assessed in order to assess the suggested model's performance. Our model is thoroughly comparing to existing sentiment analysis algorithms for these measures. The setup for conducting simulations in the proposed study is outlined in Table 2, providing comprehensive details.

Table 2: Simulation Details of used Hardware and Software.

Hardware/Software	Specification
Architecture	X86 with clock frequency of 3.4GHz, 16-cores
Processor	Inteli7,10thGen
RAM	16GB DDR4
Python	Pythonversion3.7.13(default,Apr242022,01:04:09)[GCC7.5.0]
Libraries	tensorflow , nltk, matplotlib, pandas, numpy, keras, gensim ,pickle,itertools,sys

DATASETS:

The dataset provided in this study encompasses a comprehensive collection of information pertaining to a specific subject matter. It includes a wide range of variables dataset utilized in this study is online product reviews of customer on Amazon website is used as dataset which is publically accessible on https://www.cs.jhu.edu/~mdredze/datasets/sentiment that contains 2000 (1000 negative and 1000 positive) reviews of user on each product like Book, DVD, Kitchen and Electronics. The dataset comprises various elements represented as columns. The unique identifier of the product is denoted by its ID, which is assigned a value of 8000.

- **Product Name:** The designated label for the product
- **Product Brands:** The specific brand associated with the product, such as Amazon
- **Categories:** This refers to the classification of products, such as Electronics, among others.

- **Reviews Text:** This pertains to the written evaluations provided by customers regarding a particular product.
- **Rating:** This represents the feedback provided by customers on the quality or performance of the product, typically expressed on a scale ranging from 1 to 5.

The dataset consists of a total of 8000 samples. Initially, relevant features are retrieved, whereas features exhibiting a substantial number of null values are eliminated from the dataset due to their lack of significance in predictive modelling. The ultimate dataset comprises solely of two distinct columns, namely the review text and the corresponding rating score. The rating score are categorized as Negative or Positive in the labelling process. Ratings with a value of 3 or more are classified as positive, and ratings below 3 are classified as negative.

RESULTS:

Total nine distinct permutations of word embedding and model were used for training and testing on the dataset. Each of the three deep learning models undergoes a separate phase of implementation, broken down into three sections based on the word embedding technique used.

Phase 1: In this phase, Word2Vec word embedding technique applied on the entire three deep learning models one by one as shown in Figure.

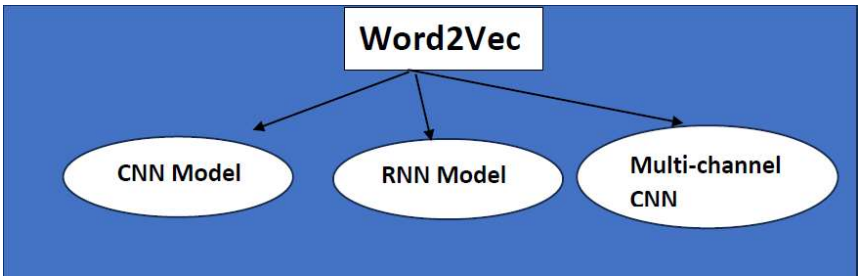

Fig 5: phase 1 model.

The table 3 shown below represents the measurement of accuracy, precision, recall and F1_score of CNN model with word2vec word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. If the kitchen customer review is the target domain and the book customer review is the source domain then the accuracy is 53%, precision is 49%, recall is 53% and F1_Score is 51%. Detailed measurement is given below.

Table 3: CNN Model using Word2Vec embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.49	0.61	0.49	0.54
2		Kitchen	0.53	0.49	0.53	0.51
3		Electronics	0.47	0.62	0.48	0.54
4	Average		0.50	0.57	0.50	0.53
5	DVD	Books	0.51	0.53	0.50	0.52
6		Kitchen	0.48	0.47	0.48	0.47
7		Electronics	0.49	0.39	0.48	0.43
8	Average		0.49	0.46	0.49	0.47
9	Kitchen	DVD	0.51	0.49	0.51	0.50
10		Books	0.48	0.47	0.48	0.47
11		Electronics	0.54	0.48	0.54	0.51
12	Average		0.51	0.48	0.51	0.49

13	Electronics	DVD	0.49	0.40	0.48	0.43
14		Kitchen	0.50	0.49	0.50	0.49
15		Books	0.51	0.28	0.51	0.36
16	Average		0.50	0.39	0.50	0.43
17	Over-all Average		0.50	0.48	0.50	0.48

The table 4 shown below represents the measurement of accuracy, precision, recall and F1_score of RNN model with word2vec word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. If an electronics customer review is used for source domain while kitchen customer review is used for target domain then the accuracy is 65%, precision is 78%, recall is 62% and F1_Score is 69%. Detailed measurement is given below in table 4.

Table 4: RNN Model using Word2Vec embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.60	0.52	0.61	0.56
2		Kitchen	0.57	0.52	0.58	0.55
3		Electronics	0.59	0.53	0.61	0.56
4	Average		0.59	0.52	0.60	0.56
5	DVD	Books	0.57	0.60	0.56	0.58
6		Kitchen	0.59	0.52	0.60	0.56
7		Electronics	0.61	0.66	0.60	0.63
8	Average		0.59	0.59	0.59	0.59
9	Kitchen	DVD	0.61	0.62	0.61	0.61
10		Books	0.61	0.49	0.65	0.56
11		Electronics	0.63	0.68	0.62	0.65
12	Average		0.62	0.60	0.63	0.61
13	Electronics	DVD	0.58	0.64	0.57	0.60
14		Kitchen	0.65	0.78	0.62	0.69
15		Books	0.61	0.78	0.58	0.66
16	Average		0.61	0.73	0.59	0.65
17	Over-all Average		0.60	0.61	0.60	0.60

The table 5 shown below represents the measurement of accuracy, precision, recall and F1_score of Multi-channel CNN model with word2vec word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. If an Electronics customer reviews is used for source domain while kitchen customer review is used for target domain then the accuracy is 74%, precision is 81%, recall is 71% and F1_Score is 76%. Detailed measurement is given below in table 5.

Table 5: Multi-channel CNN Model using Word2Vec embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.75	0.76	0.74	0.75
2		Kitchen	0.67	0.73	0.65	0.69
3		Electronics	0.66	0.71	0.65	0.68
4	Average		0.69	0.73	0.68	0.71
5	DVD	Books	0.70	0.73	0.69	0.71
6		Kitchen	0.66	0.66	0.66	0.66
7		Electronics	0.69	0.60	0.73	0.66
8	Average		0.68	0.66	0.69	0.68
9	Kitchen	DVD	0.67	0.71	0.66	0.68
10		Books	0.59	0.70	0.58	0.63

11		Electronics	0.75	0.67	0.80	0.73
12	Average		0.67	0.69	0.68	0.68
13	Electronics	DVD	0.65	0.73	0.63	0.67
14		Kitchen	0.74	0.81	0.71	0.76
15		Books	0.60	0.62	0.59	0.60
16	Average		0.66	0.72	0.65	0.68
17	Over-all Average		0.68	0.70	0.68	0.69

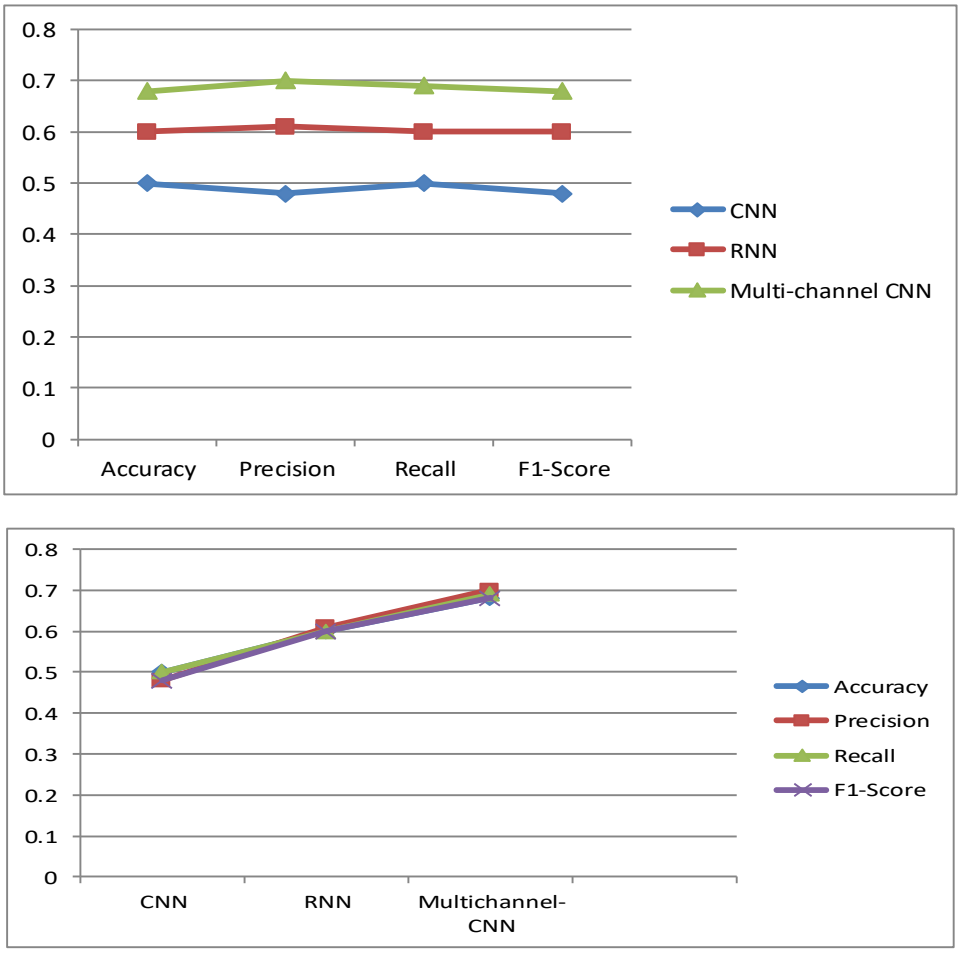

Fig: 6(a) model comparison (b) Evaluation parameter comparison

Above figure 6(a) and (b) shows the comparison between evaluation parameter like accuracy, precision, recall and F1_Score over cross domain text sentiment analysis model by using Word2Vec word embedding technique in which multi-channel CNN model gives the best result in comparison to CNN and RNN cross domain text sentiment analysis model. And figure 6(c) shows the confusion matrix of CNN, RNN and Multi-channel CNN model by using word2vec embedding techniques. Accuracy and loss of Training and validation by using word2vec word embedding in CNN RNN and multi-channel CNN respectively are shown in Figure 7(a) (b) and (c).

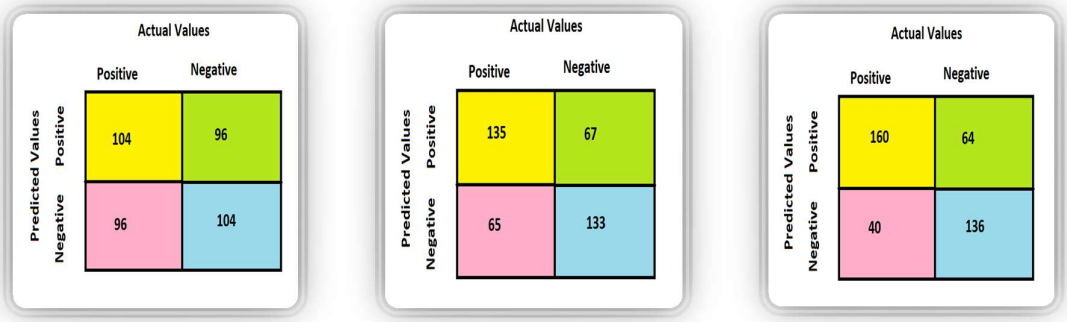

Fig: 6(c) confusion matrix

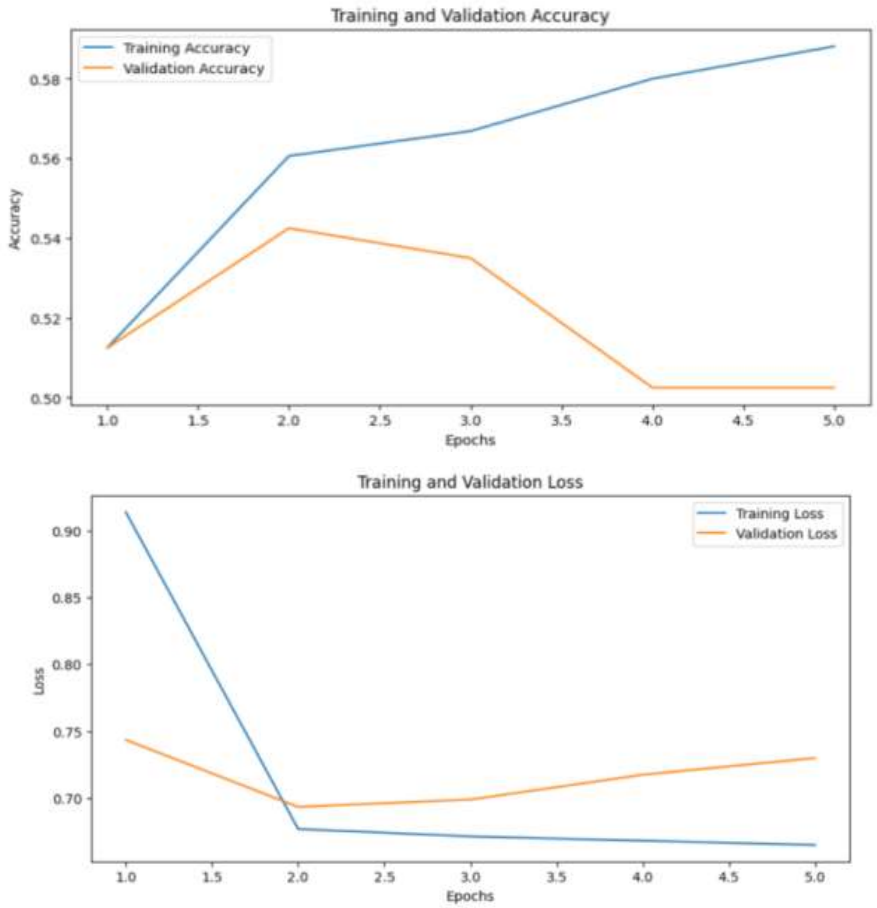

Fig 7(a): Accuracy and loss graph in CNN model

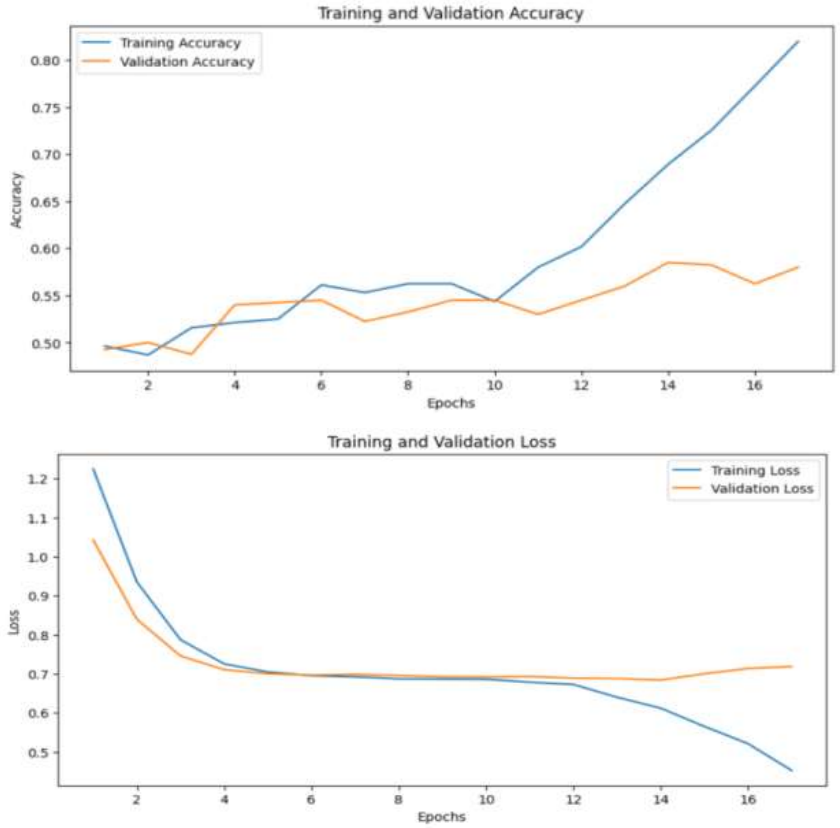

Fig 7(b): Accuracy and loss graph in RNN model

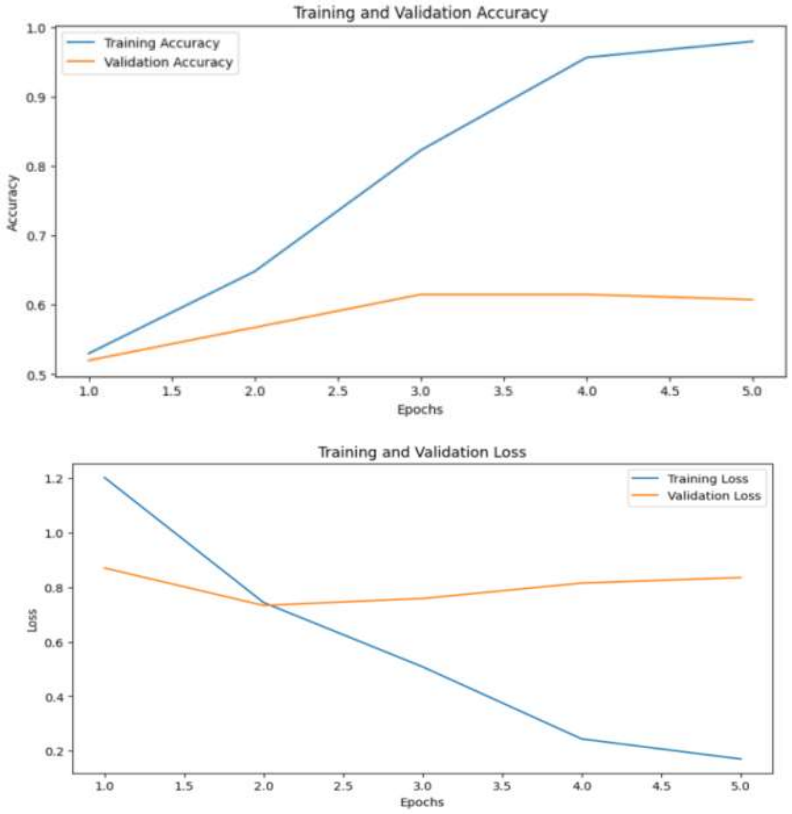

Fig 7(c): Accuracy and loss graph in multi-channel CNN model

Phase 2: In this phase, GloVe word embedding technique applied on the all three cross-domain text sentiment analysis models one by one as shown in Figure.

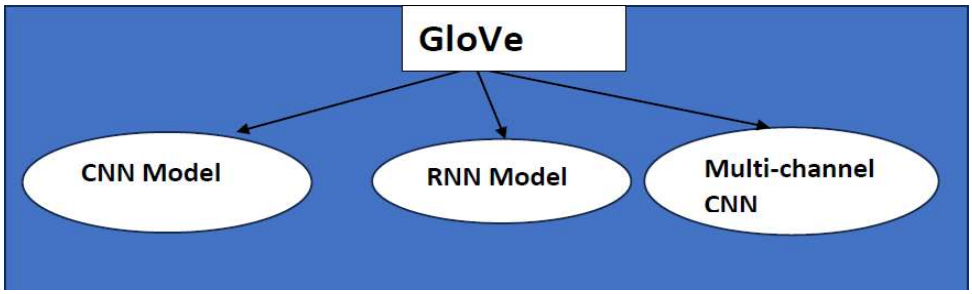

Fig 8: Phase 2 model

The table 6 shown below represents the measurement of accuracy, precision, recall and F1_score of CNN model with GloVe word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. If an Electronics customer reviews is used for source domain while kitchen customer review is used for target domain then the accuracy is 84%, precision is 84%, recall is 84% and F1_Score is 84%. Detailed measurement is given below.

Table 6: CNN Model using GloVe embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.79	0.79	0.79	0.79
2		Kitchen	0.800	0.805	0.805	0.805
3		Electronics	0.79	0.79	0.79	0.79
4	Average		0.79	0.79	0.79	0.79
5	DVD	Books	0.77	0.77	0.77	0.77
6		Kitchen	0.80	0.81	0.81	0.81
7		Electronics	0.79	0.79	0.79	0.79
8	Average		0.79	0.79	0.79	0.79
9	Kitchen	DVD	0.81	0.82	0.82	0.82
10		Books	0.82	0.82	0.82	0.82
11		Electronics	0.83	0.83	0.83	0.83
12	Average		0.82	0.82	0.82	0.82
13	Electronics	DVD	0.79	0.79	0.79	0.79
14		Kitchen	0.84	0.84	0.84	0.84
15		Books	0.78	0.78	0.78	0.78
16	Average		0.80	0.80	0.80	0.80
17	Over-all Average		0.80	0.80	0.80	0.80

The table 7 shown below represents the measurement of accuracy, precision, recall and F1_score of RNN model with GloVe word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. If an Electronics customer review is used for the source domain while kitchen customer review is used for the target domain then the accuracy is 52%, precision is 51%, recall is 51% and F1_Score is 51%. Detailed measurement is given below.

Table7: RNN Model using GloVe embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.50	0.50	0.50	0.50
2		Kitchen	0.49	0.496	0.50	0.498
3		Electronics	0.47	0.46	0.51	0.48
4	Average		0.47	0.49	0.50	0.49
5		Books	0.53	0.52	0.53	0.53

6	DVD	Kitchen	0.50	0.50	0.50	0.50
7		Electronics	0.52	0.52	0.38	0.44
8	Average		0.52	0.51	0.47	0.49
9	Kitchen	DVD	0.50	0.50	0.50	0.50
10		Books	0.50	0.50	0.50	0.50
11		Electronics	0.49	0.70	0.61	0.62
12	Average		0.50	0.57	0.54	0.54
13	Electronics	DVD	0.50	0.50	0.50	0.50
14		Kitchen	0.52	0.51	0.51	0.51
15		Books	0.51	0.51	0.51	0.51
16	Average		0.51	0.51	0.51	0.51
17	Over-all Average		0.50	0.52	0.51	0.51

The table 8 shown below represents the measurement of accuracy, precision, recall and F1_score of multi-channel CNN model with GloVe word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. If kitchen customer reviews is used for the source domain while electronics customer review is used for the target domain then the accuracy is 82%, precision is 82%, recall is 82% and F1_Score is 82%. Detailed measurement is given below.

Table 8: Multi-channel CNN Model using GloVe embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.70	0.71	0.69	0.70
2		Kitchen	0.71	0.69	0.74	0.71
3		Electronics	0.77	0.78	0.76	0.77
4	Average		0.73	0.73	0.73	0.73
5	DVD	Books	0.73	0.73	0.73	0.73
6		Kitchen	0.59	0.61	0.55	0.58
7		Electronics	0.74	0.75	0.74	0.75
8	Average		0.69	0.70	0.67	0.69
9	Kitchen	DVD	0.80	0.80	0.80	0.80
10		Books	0.74	0.72	0.77	0.74
11		Electronics	0.82	0.82	0.82	0.82
12	Average		0.79	0.78	0.80	0.79
13	Electronics	DVD	0.73	0.73	0.73	0.73
14		Kitchen	0.77	0.77	0.78	0.78
15		Books	0.75	0.75	0.75	0.75
16	Average		0.75	0.75	0.76	0.76
17	Over-all Average		0.74	0.74	0.74	0.74

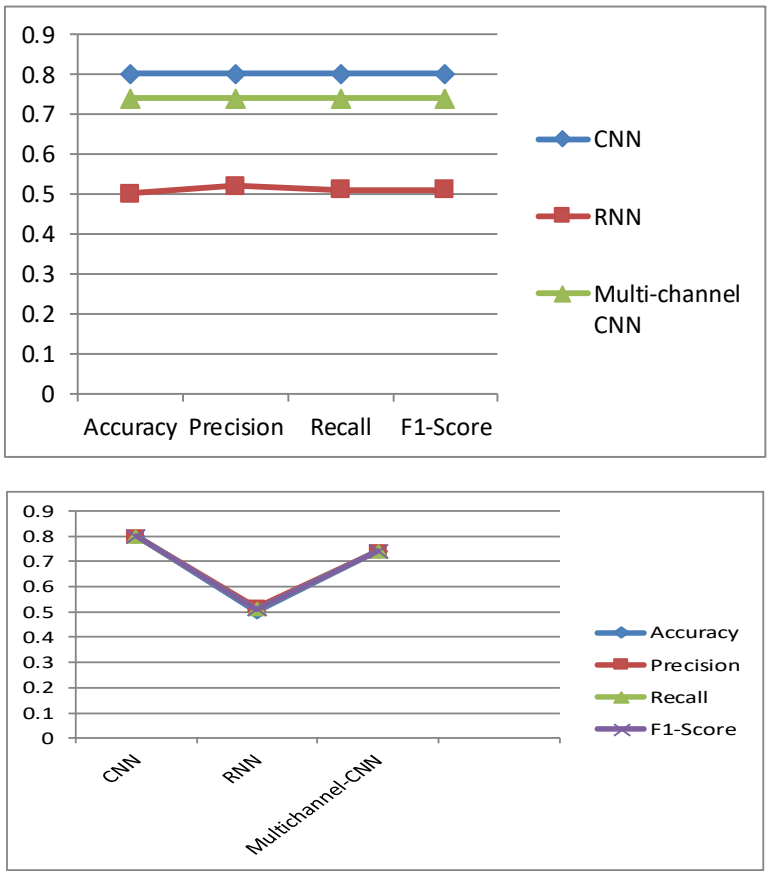

Fig: 9(a) Model comparison (b) parameter comparison

Figure 9 (a) (b) and (c) shows the comparison between evaluation parameter like accuracy, precision, recall and F1_Score and confusion matrix over cross domain text sentiment analysis model by using GloVe word embedding technique in which CNN model gives the best result in comparison to multi-channel CNN, and RNN cross domain text sentiment analysis model. Accuracy and loss of Training and validation by using GloVe word embedding in CNN RNN and multi-channel CNN respectively are shown in Figure 10 (a) (b) and (c).

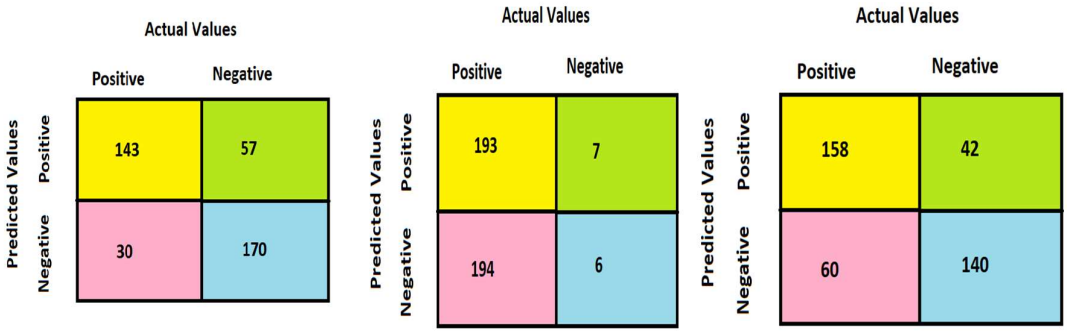

Figure 9(c): Confusion Matrix

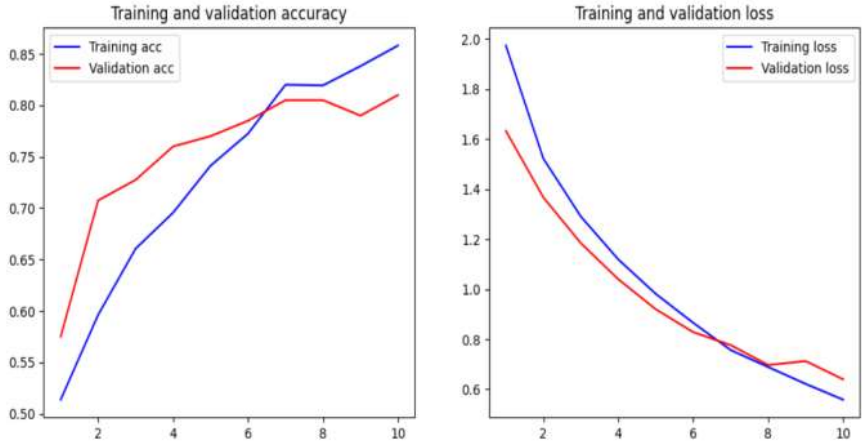

Fig 10(a): Accuracy and loss graph in CNN model

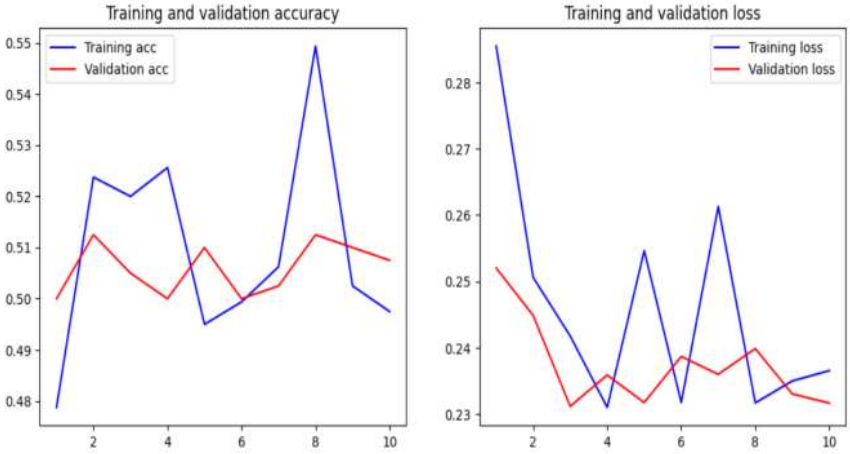

Fig 10(b): Accuracy and loss graph in RNN model

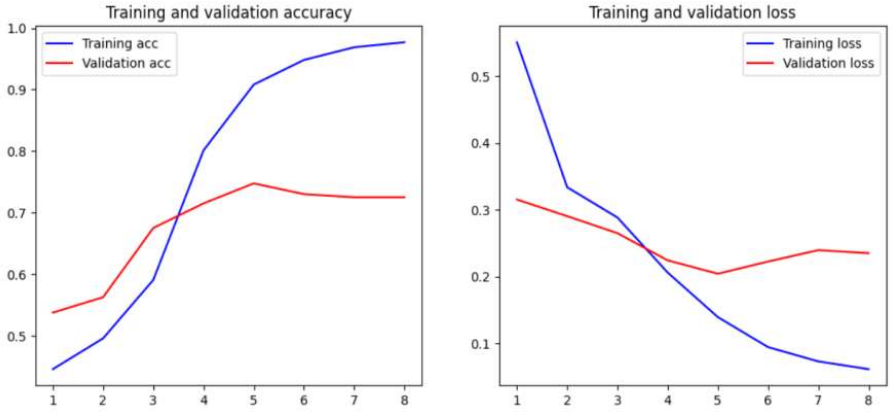

Fig 10(c): Accuracy and loss graph in multi-channel CNN model

Phase 3: In this phase, GloVe word embedding technique applied on the entire three deep learning models one by one as shown in Figure 11.

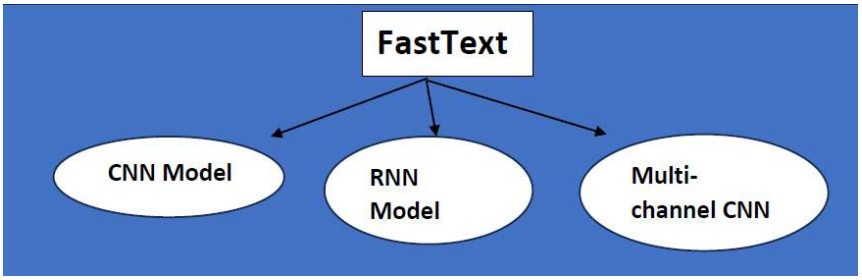

Fig 11: Phase 3 model

The table 10 shown below represents the measurement of various parameters such as accuracy, precision, recall and F1_score of CNN model using FastText word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. For instance, if customer reviews of kitchen dataset is used for the source domain while customer reviews of electronics dataset is used for target domain, then the accuracy is 60%, precision is 68%, recall is 59% and F1_Score is 63%. As like, remaining combinations of source domain and destination domain are listed below with detailed measurement values.

Table9: CNN Model using FastText embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.50	0.45	0.50	0.47
2		Kitchen	0.51	0.49	0.51	0.50
3		Electronics	0.48	0.53	0.48	0.50
4	Average		0.50	0.49	0.50	0.49
5	DVD	Books	0.56	0.45	0.58	0.50
6		Kitchen	0.58	0.49	0.59	0.54
7		Electronics	0.53	0.37	0.54	0.44
8	Average		0.56	0.44	0.57	0.49
9	Kitchen	DVD	0.52	0.55	0.52	0.53
10		Books	0.51	0.49	0.51	0.50
11		Electronics	0.60	0.68	0.59	0.63
12	Average		0.54	0.57	0.54	0.55
13	Electronics	DVD	0.53	0.55	0.52	0.53
14		Kitchen	0.63	0.64	0.63	0.63
15		Books	0.52	0.55	0.52	0.53
16	Average		0.56	0.58	0.56	0.56
17	Over-all Average		0.54	0.52	0.54	0.52

The table 10 shown below represents the measurement of various parameters such as accuracy, precision, recall and F1_score of RNN model using FastText word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. For instance, if customer reviews of kitchen dataset is used for source domain while customer reviews of electronics dataset is used for target domain, then the accuracy is 76%, precision is 65%, recall is 83% and F1_Score is 73%. As like, remaining combinations of source domain and destination domain are listed below with detailed measurement values.

Table10: RNN Model using FastText embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.68	0.73	0.67	0.70
2		Kitchen	0.68	0.73	0.68	0.70
3		Electronics	0.56	0.50	0.56	0.53
4	Average		0.64	0.65	0.64	0.64
5		Books	0.60	0.72	0.58	0.64

6	DVD	Kitchen	0.65	0.65	0.65	0.65
7		Electronics	0.63	0.55	0.65	0.60
8	Average		0.63	0.64	0.63	0.63
9	Kitchen	DVD	0.65	0.76	0.62	0.68
10		Books	0.60	0.60	0.60	0.60
11		Electronics	0.76	0.65	0.83	0.73
12	Average		0.67	0.67	0.68	0.67
13	Electronics	DVD	0.66	0.66	0.66	0.67
14		Kitchen	0.75	0.72	0.76	0.74
15		Books	0.64	0.69	0.63	0.66
16	Average		0.68	0.69	0.68	0.69
17	Over-all Average		0.66	0.66	0.66	0.66

The table 11 shown below represents the measurement of various parameters such as accuracy, precision, recall and F1_score of multi-channel CNN model using FastText word embedding technique. It covers all the possibilities of source and destination domain of given four amazon product review such as Book, DVD, Kitchen and Electronics. For instance, if customer reviews of kitchen dataset is used for the source domain while customer review of electronics dataset is used for the target domain, then the accuracy is 76%, precision is 66%, recall is 82% and F1_Score is 73%. As like, remaining combinations of source domain and destination domain are listed below with detailed measurement values.

Table11: Multi-channel CNN Model using FastText embedding techniques.

S.N	Source Domain	Destination Domain	Accuracy	Precision	Recall	F1_Score
1	Books	DVD	0.74	0.80	0.71	0.75
2		Kitchen	0.67	0.68	0.66	0.67
3		Electronics	0.66	0.65	0.67	0.66
4	Average		0.69	0.71	0.68	0.69
5	DVD	Books	0.70	0.66	0.71	0.68
6		Kitchen	0.68	0.64	0.70	0.66
7		Electronics	0.71	0.62	0.75	0.68
8	Average		0.70	0.64	0.72	0.67
9	Kitchen	DVD	0.67	0.72	0.66	0.68
10		Books	0.63	0.72	0.61	0.66
11		Electronics	0.76	0.66	0.82	0.73
12	Average		0.69	0.70	0.70	0.69
13	Electronics	DVD	0.63	0.70	0.62	0.65
14		Kitchen	0.73	0.80	0.71	0.75
15		Books	0.62	0.63	0.62	0.63
16	Average		0.66	0.71	0.65	0.68
17	Over-all Average		0.69	0.69	0.69	0.68

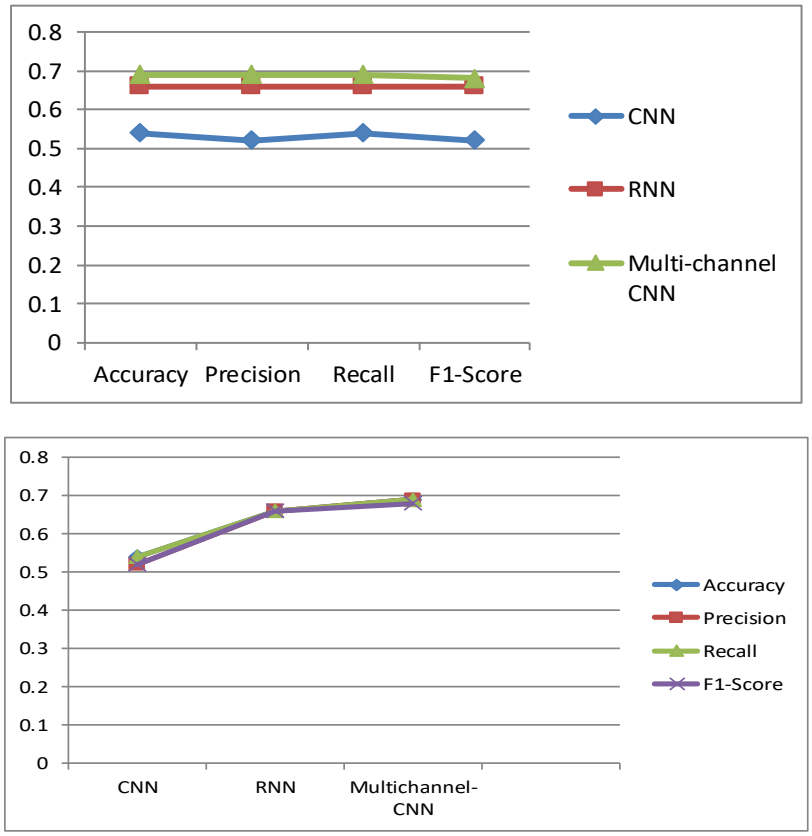

Fig: 12(a) Model comparison (b) Parameter comparison

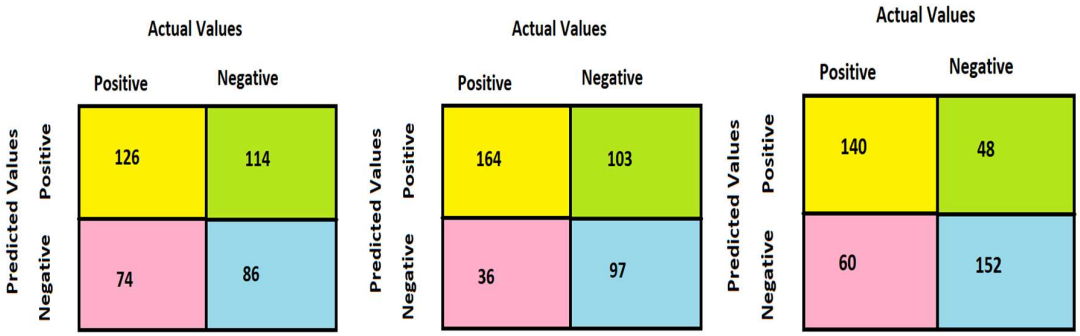

Fig: 12(c) Confusion matrix

In the above figure 12 (a) (b) and (c) shows the comparison between evaluation parameter like accuracy, precision, recall and F1_Score and confusion matrix over cross domain text sentiment analysis model by using FastText word embedding technique in which multi-channel CNN model gives the best result in comparison to CNN and RNN cross domain text sentiment analysis model. Accuracy and loss of training and validation by using FastText word embedding in CNN RNN and multi-channel CNN respectively are shown in Figure 13(a) (b) and (c).

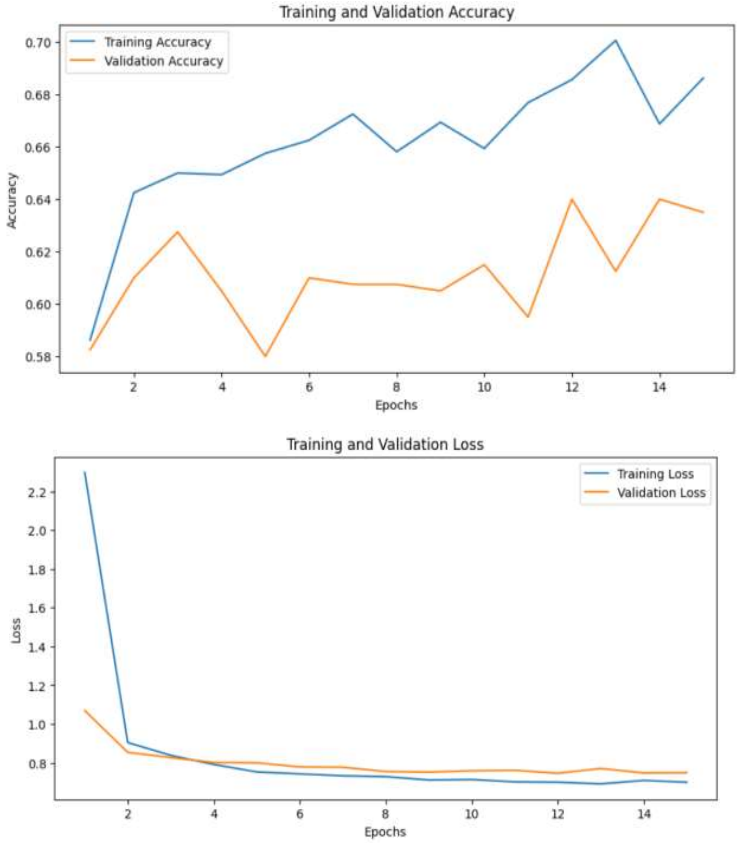

Fig 13(a): Accuracy and loss graph in CNN model

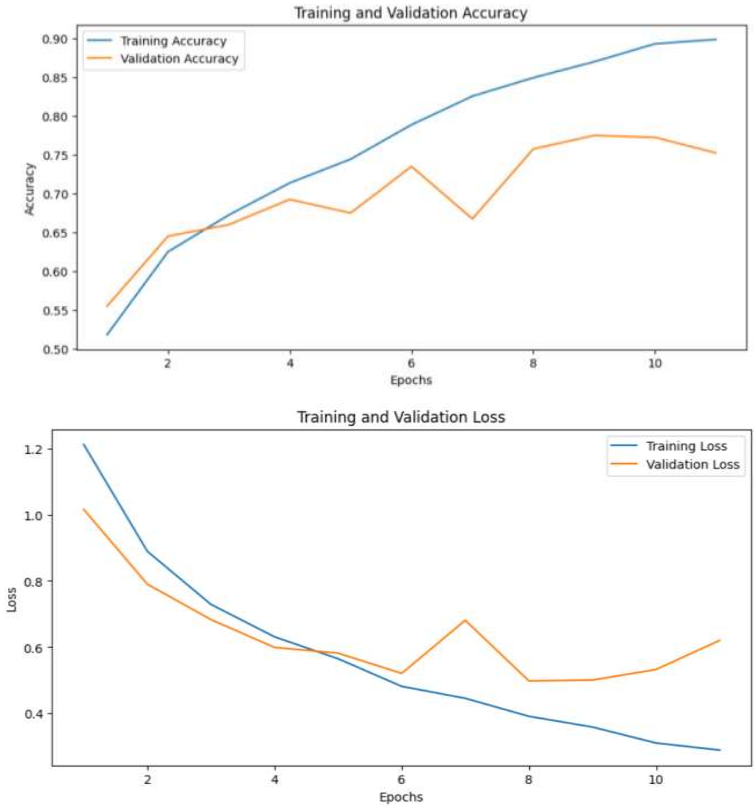

Fig 13 (b): Accuracy and loss graph in RNN model

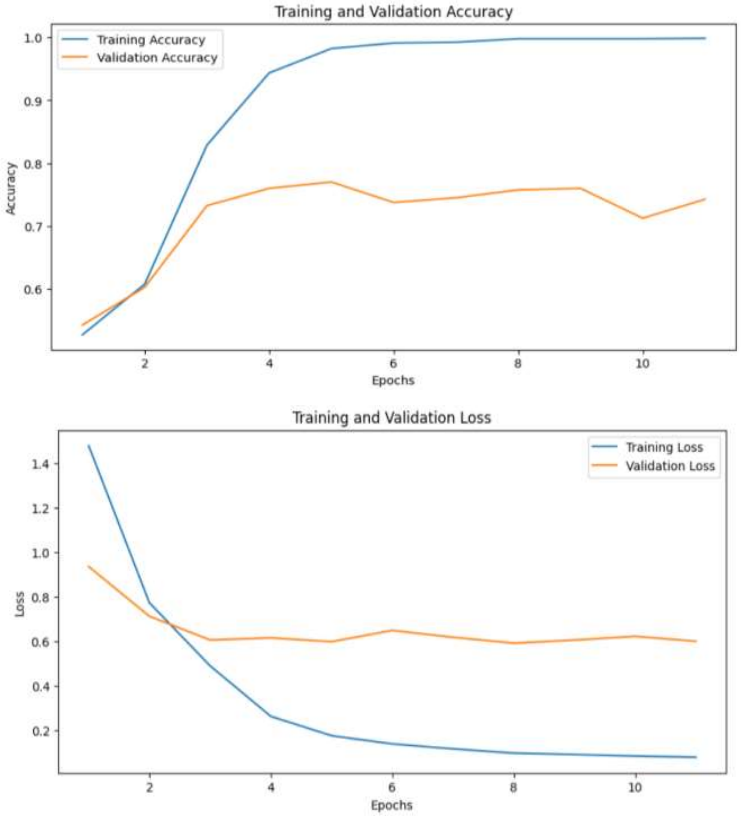

Fig 13 (c): Accuracy and loss graph in multi-channel CNN model

This article focuses on creating an impartial and efficient method that will use the neutrality of reviews to improve the analysis.

- Using word2vec word embedding technique, multichannel CNN model produced maximum accuracy of 75% among all models.
- Using gloVe word embedding technique, CNN model produced maximum accuracy of 84% among all models.
- Using fastText word embedding technique, multi-channel CNN model produced maximum accuracy of 75% among all models.

Table12: Overall Result

S.N	Word Embedding	Model	Source Domain	Destination Domain	Max-Accuracy Obtained
1	Word2Vec	CNN	Book	Kitchen	53%
2		RNN	Book	DVD	60%
3		Multi-Channel CNN	Book	DVD	75%
4	GloVe	CNN	Electronics	Kitchen	84%
5		RNN	DVD	Book	53%
6		Multi-Channel CNN	Kitchen	Electronics	82%
7	FastText	CNN	Electronics	Kitchen	63%
8		RNN	Kitchen	Electronics	76%
9		Multi-Channel CNN	Kitchen	Electronics	76%

6. CONCLUSION

In this paper, we tested three popular supervised classifiers on the amazon review dataset for analysing customer reviews. The results revealed that proposed model could perform cross domain sentiment analysis with an accuracy of 84% when the training is done on electronics data set and testing on kitchen data set by using glove word embedding techniques. The Multi-channel CNN model using the GloVe word embedding technique provides the highest accuracy rate of any of the models at around 84%, followed by the CNN model at 82%.The paper's results demonstrate the model's superiority, with multi-channel CNN with GloVe word embedding achieving up to 84% accuracy when compared to SVM, LSTM-CNN, and ParsBERT models [56]. When comments used in training and testing the model from the same domain but from different time periods, ParsBERT models achieved 82% SVM 76% and LSTM-CNN 58% accuracy respectively. In more complicated scenarios, ParsBERT models achieved 68% SVM 58% and LSTM-CNN 55% accuracy when comments used in training and testing the model from both different domains and different time periods.The positive and negative separation of comments is used to analyse the quality of the online products. Cross Domain Sentiment Analysis achieves relatively good performance. From our analysis on amazon product reviews data set, CNN model works best for cross domain text sentimental analysis when performed word embedding using GloVe. Accuracy and other performance metric could be improved by proposing hybrid model using CNN and RNN.

REFERENCES

[1] N. Hajli, ''Social commerce constructs and consumer's intention to buy,''Int. J. Inf. Manage., vol. 35, no. 2, pp. 183–191, 2015.

[2] V. Jha, R. Savitha, P. D. Shenoy, K. Venugopal, and A. K. Sangaiah,''A novel sentiment aware dictionary for multi-domain sentiment classification,'' Comput. Elect. Eng., vol. 69, pp. 585–597, Jul. 2018.

[3] L. Peng, G. Cui, M. Zhuang, and C. Li. (2014). What Do Seller Manipulations of Online Product Reviews Mean to Consumers.Accessed: Jul. 2019.[Online]. Available: http://commons.ln.edu.hk/hkibswp/70.

[4] P. Ducange, M. Fazzolari, M. Petrocchi, and M. Vecchio, ''An effectiveDecision Support System for social media listening based on cross-sourcesentiment analysis models,'' Eng. Appl. Artif. Intell., vol. 78, pp. 71–85,Feb. 2019.

[5] A. B. Eliacik and N. Erdogan, ''Influential user weighted sentiment analysis on topic based microblogging community,'' Expert Syst. With Appl.,vol. 92, pp. 403–418, Feb. 2018.

[6] A. Liu and R. Ireland ''Application of data analytics for product design:Sentiment analysis of online product reviews,'' CIRP J. Manuf. Sci. Technol., vol. 23, pp. 128–144, Nov. 2018.

[7] B. Heredia, T. M. Khoshgoftaar, J. Prusa, and M. Crawford, ''Integratingmultiple data sources to enhance sentiment prediction,'' in Proc. IEEE 2nd Int. Conf. Collaboration Internet Comput. (CIC), Pittsburgh, PA, USA,Nov. 2016, pp. 285–291.

[8] F. Jin, W. Wang, P. Chakraborty, N. Self, F. Chen, and N. Ramakrishnan,''Tracking multiple social media for stock market event prediction,'' inProc. Ind. Conf. Data Mining, Ney York, NY, USA, 2017, pp. 16–30.

[9] D. Bollegala, D. Weir, and J. Carroll, ''Cross-domain sentiment classification using a sentiment sensitive thesaurus,'' IEEE Trans. Knowl. DataEng., vol. 25, no. 8, pp. 1719–1731, Aug. 2013.

[10] A. Tsakalidis, S. Papadopoulos, and I. Kompatsiaris, ''An ensemble modelfor cross-domain polarity classification on twitter,'' in Proc. Int. Conf. WebInf. Syst. Eng., Thessaloniki, Greece, 2014, pp. 168–177.

[11] Y. Zhang, X. Hu, P. Li, L. Li, and X. Wu, ''Cross-domain sentimentclassification-feature

divergence, polarity divergence or both?'' PatternRecognit.Lett., vol. 65, pp. 44–50, Nov. 2015.

[12] G. Zhou, Y. Zhou, X. Guo, X. Tu, and T. He, ''Cross-domain sentiment classification via topical correspondence transfer,'' Neurocomputing,vol. 159, pp. 298–305, Jul. 2015.

[13] H. Hammer, A. Yazidi, A. Bai, and P. Engelstad, ''Building domain specificsentiment lexicons combining information from many sentiment lexiconsand a domain specific corpus,'' in Proc. IFIP Int. Conf. Comput. Sci. ItsAppl., Saida, Algeria, 2015, pp. 205–216.

[14] M. Dragoni and G. Petrucci, ''A neural word embeddings approach formulti-domain sentiment analysis,'' IEEE Trans. Affect. Comput., vol. 8,no. 4, pp. 457–470, Oct./Dec. 2017.

[15] J. Blitzer, R. McDonald, and F. Pereira, ''Domain adaptation with structural correspondence learning,'' in Proc. Conf. Empirical Methods Natural Lang. Process. Assoc. Comput. Linguistics, Sydney, NSW, Australia, 2006, pp. 120–128.

[16] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, ''Cross-domain sentiment classification via spectral feature alignment,'' in Proc. 19th Int. Conf. World Wide Web, Raleigh, NC, USA, 2010, pp. 751–760.

[17] F. Wu and Y. Huang, ''Sentiment domain adaptation with multiple sources,'' in Proc. 54th Annu.Meeting Assoc. Comput. Linguistics, Berlin, Germany, 2016, pp. 301–310.

[18] F. Wu, S. Wu, Y. Huang, S. Huang, and Y. Qin, ''Sentiment domain adaptation with multi-level contextual sentiment knowledge,'' in Proc. 25th ACM Int. Conf. Inf. Knowl.Manage., Indianapolis, IN, USA, 2016, pp. 949–958.

[19] Z. Yuan, S. Wu, F. Wu, J. Liu, and Y. Huang, ''Domain attention model for multi-domain sentiment classification,'' Knowl.-Based Syst., vol. 155, pp. 1–10, Sep. 2018.

[20] K. Ravi and V. Ravi, ''A survey on opinion mining and sentiment analysis: Tasks, approaches and applications,'' Knowl.-Based Syst., vol. 89, pp. 14–46, Nov. 2015.

[21] T. Al-Moslmi, N. Omar, S. Abdullah, and M. Albared, ''Approaches to cross-domain sentiment analysis: A systematic literature review,'' IEEE Access, vol. 5, pp. 16173–16192, 2017.

[22] G. Katz, N. Ofek, and B. Shapira, ''ConSent: Context-based sentiment analysis,'' Knowl.-Based Syst., vol. 84, pp. 162–178, Aug. 2015.

[23] Kongthon, A., Sangkeettrakarn, C., Kongyoung, S. & Haruechaiyasak, C. Implementing an online help desk system based on conversational agent. In Proceedings of the International Conference on Management of Emergent Digital EcoSystems 2009 Oct 27 (pp. 450–451).

[24] Jean, S., Cho, K., Memisevic, R. & Bengio, Y. On using very large target vocabulary for neural machine translation.arXiv preprintarXiv:1412.2007.2014Dec5.

[25] Liang, B., Hang, S., Gui, L., Cambria, E. & Ruifeng, X. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. Knowledge Based Syst. 235, 107643 (2022).

[26] Strubell,E.,Ganesh,A.&McCallum,A.EnergyandpolicyconsiderationsfordeeplearninginNLP.arXiv: 1906.02243.2019Jun5.

[27] Fisher,I.E.,Garnsey,M.R.&Hughes,M.E.Naturallanguageprocessinginaccounting,auditingandfina nce:Asynthesisoftheliteraturewitharoadmapfor future esearch.Intell.Syst.Account.FinanceManag.23(3),157–214(2016).

[28] Vinyals,O.,Fortunato,M.&Jaitly,N.Pointernetworks.Advancesinneuralinformationprocessingsyste ms.2015;

[29] Mnasri,M.RecentadvancesinconversationalNLP:TowardsthestandardizationofChatbotbuilding.ar

Xiv:1903.09025.2019Mar21.

[30] Moreno,A.&Redondo,T.Textanalytics:theconvergenceofbigdataandartificialintelligence.IJIMAI,57–64(2016).

[31] Srinivasu,P.N.,Bhoi,A.K.,Jhaveri,R.H.,Reddy,G.T.&Bilal,M.ProbabilisticDeepQNetworkforreal-timepathplanningincensoriousroboticproceduresusingforcesensors.J.RealTimeImageProcess.18(5),1773–1785(2021).

[32] Kumar, A., Abhishek, K., Chakraborty, C. & Kryvinska, N. Deep learning and internet of things based lung ailment recognitionthroughcoughingspectrograms.IEEEAccess.1(9),95938–48(2021).

[33] Khamparia, A. et al. Internet of health things-driven deep learning system for detection and classification of cervical cells usingtransferlearning.J.Supercomput.76(11),8590–8608(2020).

[34] ECambria,Qliu,Sdecherchi,Fxing,Kkwok.Senticnet:acommonsensebasedneurosymbolicaiframeworkforexplainablesentimentanalysis.Proceedingsofthe13thConferenceonLanguageResourcesandEvaluation(LREC2022):3829-3839.

[35] Trueman,T.E.&Cambria,E.Aconvolutionalstackedbidirectional LSTMwithamultiplicativeattention mechanismforaspectcategoryand sentiment etection.Cogn.Comput.13(6),1423–1432(2021).

[36] Cambria,Erik,DipankarDas,SivajiBandyopadhyay,andAntonioFeraco.Affectivecomputingandsentimentanalysis.InApracticalguidetosentimentanalysis,pp.1-10.Springer,Cham,2017.

[37] He,Kai,RuiMao,TieliangGong,ChenLi,andErikCambria. Meta-based Self-training and Re-weightingforAspect-based SentimentAnalysis.IEEETransactionsonAffectiveComputing(2022).

[38] Mao,Rui,QianLiu,KaiHe,WeiLi,andErikCambria.Thebiasesofpre-trainedlanguagemodels:Anempiricalstudyonprompt-basedsentimentanalysisandemotiondetection.IEEETransactionsonAffectiveComputing(2022).

[39] Sharf, Z.; Rahman, S.U. Performing natural language processing on roman urdu datasets. Int. J. Comput. Sci. Netw. Secur.**2018**, 141–148.

[40] Javed, I.; Afzal, H. Creation of bi-lingual social network dataset using classifiers. In International Workshop on Machine Learningand Data Mining in Pattern Recognition; Springer: Cham, Switzerland, 2014; pp. 523–533.

[41] Mehmood, F.; Ghani, M.U.; Ibrahim, M.A.; Shahzadi, R.; Mahmood,W.; Asim, M.N. A precisely xtreme-multi channel hybridapproach for roman urdu sentiment analysis. IEEE Access **2020**, 8, 192740–192759. [CrossRef]

[42] Mahmood, Z.; Safder, I.; Nawab, R.M.A.; Bukhari, F.; Nawaz, R.; Alfakeeh, A.S.; Aljohani, N.R.; Hassan, S.U. Deep sentiments inroman urdu text using recurrent convolutional neural network model. Inf. Process. Manag.**2020**, 57, 102233.

[43] Mehmood, K.; Essam, D.; Shafi, K.; Malik, M.K. Sentiment analysis for a resource poor language—Roman Urdu. ACM Trans.Asian-Low-Resour. Lang. Inf. Process. (TALLIP) **2019**, 19, 1–15.

[44] Mehmood, K.; Essam, D.; Shafi, K.; Malik, M.K. An unsupervised lexical normalization for Roman Hindi and Urdu sentimentanalysis. Inf. Process. Manag.**2020**, 57, 102368.

[45] Khan, L.; Amjad, A.; Afaq, K.M.; Chang, H.-T. Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media. *Appl. Sci.* **2022**, *12*, 2694.https://doi.org/10.3390/app12052694

[46] Meng, J.N.; Yu, Y.H.; Zhao, D.D.; Sun, S. Cross-domain sentiment analysis based on combination

of featureand instance-transfer. J. Chin. Inf. Process. **2015**, 79, 74–79, 143.

[47] Huang, R.Y.; Kang, S.Z. Improved EM-based cross-domain sentiment classification method. Appl. Res.Comput. **2017**, 34, 2696–2699.

[48] Xia, R.; Zong, C.Q.; Hu, X.L.; Cambria, E. Feature ensemble plus sample selection: Domain adaptation forsentiment classification. IEEE Intell. Syst. **2013**, 28, 10–18. [CrossRef]

[49] Tang, D.Y.; Qin, B.; Liu., T. Deep learning for sentiment analysis: Successful approaches and future challenges.WileyInterdiscip. Rev. Data Min. Knowl. Discov.**2015**, 5, 292–303. [CrossRef]

[50] Yu, J.F.; Jiang, J. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification.In Proceedings of the Conference on Empirical Methods in Natural Language, Stroudsburg, PA, USA, 1–4November 2016.

[51] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[52] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532-1543. 2014.

[53] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.

[54] Meng, Jiana, Yingchun Long, Yuhai Yu, Dandan Zhao, and Shuang Liu. "Cross-domain text sentiment analysis based on CNN_FT method." *Information* 10, no. 5 (2019): 162.

[55] Y. Fan, X. Mi and Y. Nie, "Cross-Domain Discriminative Subspace Classification Algorithm for Review Text Sentiment Recognition Oriented E-Commerce Platforms," in *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 3455-3463, Feb. 2024, doi: 10.1109/TCE.2024.3372503.

[56] Panahandeh Nigjeh, Mahnaz, and ShirinGhanbari."Leveraging ParsBERT for cross-domain polarity sentiment classification of Persian social media comments." Multimedia Tools and Applications 83, no. 4 (2024): 10677-10694.

[57] Singh, U., Saraswat, A., Azad, H.K. *et al.* Towards improving e-commerce customer review analysis for sentiment detection. *Sci Rep* **12**, 21983 (2022).https://doi.org/10.1038/s41598-022-26432-3