PREDICTING OVERALL CUSTOMER SATISFACTION FOR AN EFFECTIVE OPINION ANALYSIS ON E-COMMERCE WEBSITES

ई कॉमर्स वेबसाइट के लिए प्रभावी ग्राहक जज्बात निष्कर्ष प्रणाली

A

Thesis

Submitted for the Award of the Ph.D. degree of

PACIFIC ACADEMY OF HIGHER EDUCATION AND RESEARCH UNIVERSITY

By

VIJAYA RAVINDRA SAGVEKAR

विजया रविन्द्र सागवेकर

Under the supervision of

Dr. PRASHANT SHARMA

Associate Professor,
Pacific Academy of Higher
Education & Research University, Udaipur

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER ENGINEERING
PACIFIC ACADEMY OF HIGHER EDUCATION AND RESEARCH UNIVERSITY, UDAIPUR

2023

DECLARATION

I, **VIJAYA RAVINDRA SAGVEKAR D/O MR. RAVINDRA WASUDEO SAGVEKAR** resident of B-3, Vaishali CHLS, Near Teacher's colony, Bandra East, Mumbai, State- Maharashtra, Pincode-400051, hereby declare that the research work incorporated in the present thesis entitled **"Predicting Overall Customer Satisfaction For An Effective Opinion Analysis On E-Commerce Websites"** (ई कॉमर्स वेबसाइट के लिए प्रभावी ग्राहक जज्बात निष्कर्ष प्रणाली) is my original work. This work (in part or in full) has not been submitted to any University for the award or a Degree or a Diploma. I have properly acknowledged the material collected from secondary sources wherever required.

I solely own the responsibility for the originality of the entire content.

Signature of the Candidate

Date:

FACULTY OF ENGINEERING

PACIFIC ACADEMY OF HIGHER EDUCATION AND RESEARCH UNIVERSITY, UDAIPUR

Dr. PRASHANT SHARMA
Associate Professor

CERTIFICATE

It gives me immense pleasure in certifying that the thesis **"Predicting Overall Customer Satisfaction For An Effective Opinion Analysis On E-Commerce Websites"** (ई कॉमर्स वेबसाइट के लिए प्रभावी ग्राहक जज्बात निष्कर्ष प्रणाली) and submitted by **VIJAYA RAVINDRA SAGVEKAR** is based on the research work carried out under my guidance. He / she have completed the following requirements as per Ph.D. regulations of the University;

(i) Course work as per the University rules.

(ii) Residential requirements of the University.

(iii) Regularly presented Half Yearly Progress Report as prescribed by the University.

(iv) Published / accepted minimum of two research paper in a refereed research journal.

I recommend the submission of thesis as prescribed/notified by the University.

Date: **Name and Designation of Supervisor**

Dr. PRASHANT SHARMA
Associate Professor,
Pacific Academy of Higher Education
& Research University, Udaipur

COPYRIGHT

I, **VIJAYA RAVINDRA SAGVEKAR,** hereby declare that the Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan, shall have the rights to preserve, use and disseminate this dissertation entitled **"Predicting Overall Customer Satisfaction For An Effective Opinion Analysis On E-Commerce Websites" (ई कॉमर्स** वेबसाइट के लिए प्रभावी ग्राहक जज्बात निष्कर्ष प्रणाली) in print or in electronic format for the academic research.

Date:

Signature of Candidate

Place:

ACKNOWLEDGEMENT

Foremost, I am highly indebted to **Dr. Prashant Sharma**, Associate Professor, Pacific Academy of Higher Education & Research University for his guidance and constant supervision as well as for providing necessary information regarding the thesis work and also for his support in completing the thesis work. I would like to express my sincere gratitude towards him for his continuous support of my Ph.D. study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank my colleague **Dr. Pradip Suresh Mane** for his encouragement and insightful comments. Special thanks to Pacific staff **Mrs. Kusum Sharma.**

Last but not the least; I would like to thank my family for supporting me spiritually throughout my life. My special thanks to my husband **Mr. Anuj Jyant Bijoor** for his support and help in all ways I needed. My father **Late. Shri. Ravindra Wasudeo Sagvekar** whose blessings are always with me, last but not the least, I am thankful to my mother **Shrimati Shobha Sagvekar** , mother-in-law **Shrimati**. **Jyoti Jayant Bijoor**, sisters **Mrs.Vrushali Devrukkhar**, **Mrs**. **Kalpana Deorukhkar** and **Mrs. Vidya Sagvekar-Desai**.

I am very grateful to **Dr. Surya Prakash Vaishnav**, Assistant Professor, PAHER University and friend for his valuable suggestions and encouragement throughout the period of my research.

Last but not the least, my distinctive thanks to **Nav Nimantran, (Hemant Sharma, Kusum)** Udaipur for their role in shaping the matter, creative design work and bringing out this document meticulously, neatly and timely.

DATE: - **VIJAYA RAVINDRA SAGVEKAR**

DEDICATED TO

MY FAMILY, FRIENDS

AND WELL-WISHERS

ABSTRACT

Online stores now collect a large amount of client feedback in the form of surveys, reviews, and comments. This feedback is categorized in some circumstances for often underused - despite the fact that customer choice is critical to their business's success. E-commerce websites have evolved in a wide variety of the advantages of marketing for consumers to publish or share their engagement with the obtained goods by writing reviews that include beneficial remarks, thoughts, and product review. Nowadays, a huge quantity of customers have the ability to compare products in digital stores as well as select their topmost selections in automated merchants, for example, Amazon.com and Taobao.com.Sentiment Analysis is widely used as the voice of the client in applications aimed towards displaying and client care. Sentiment extractors, in their top basic framework, categorize communications as having a positive, negative, or occasionally neutral premise. User reviews on e-commerce websites give useful details about the item. Sentiment analysis on the text review contributions in analysing the sentiment of users about the product and predicting product sales. Existing sentiment analysis approaches include Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN)-based methods, which suffer from the threatened gradient and overfitting problems.

In the first study, suggested a classification model based on deep learning techniques for determining the review condition. The findings indicated that the recommended techniques for the client were based on previous reviews, initial assessments, and the solutions provided in response to the client's inquiry audit.Furthermore, it showed the suggested techniques was organized to classify each of the reviews exhibits a remarkable similarity, resembling a human response to the customer. In the second study,proposed Word Embedding Attention (WEA) technique is proposed in Bi-directional Long Short Term Memory (Bi-LSTM) method for increases the performance of the classification. The CNN model is applied to identify the attributes from the input dataset for sentiment analysis. The Balanced Cross-Entropy is proposed to maintain the gradient in the network and solves vanishing gradient problem in the network. The WEA technique provides higher weight to the words having strong relation with class. This technique helps to increases the performance of the model related to class-wise, thus increases the precision and recall value. The CNN feature helps to provide higher performance for less number of training data. The WEA-Bi-LSTM model has 97.4 % accuracy, 86.8 % precision and existing CNN model has 97.1 % accuracy, and 85.4 % precision in sentiment analysis.

LIST OF CONTENTS

LIST OF TABLE

LIST OF FIGURE

LIST OF ABBREVIATIONS

Aspect Level or Aspect Based Sentiment Analysis	ABSA
Bag of Words	BOW
Collaborative Filtering	CF
Content-Based	CB
Convolutional Neural Network	CNN
Convolutional Neural Network	CNN
Deep Learning	DL
Fake Review Detection Framework	FRDF
Gated Recurrent Unit	GRU
Gensim Lemmatization	GL
Hybrid Recommendation System	HRS
Latent Dirichlet Allocation	LDA
Local Search Improvised Bat Algorithm based Elman Neural Network	LSIBA-ENN
Long Short Term Memory	LSTM
Machine Learning	ML
Machine Learning Classifiers	MLC
Naive Bayes	NB
Natural Language Processing	NLP
Pointwise Mutual Information	PMI
Product Comment Summarizer and Analyzer	PCSA
Snow-Ball Stemming	SBS
Support Vector Machines	SVM
Transfer learning	TL
User-Generated Content	UGC
User-Generated Content	UGC
Waikato Environment for Knowledge Analysis	WEKA
Web Scrapping Tool	WST
Word Embedding Attention	WEA
World Wide Web	WWW

CHAPTER – I

INTRODUCTION

1.1 Overview

In recent times, a greater number of customers wish to buy their items over the internet instead of physically in order to save time and obtain a good deal on the items [1]. Mostly every customer reads reviews over specific items or comparable kinds of items before buying them internet [2]. Customers communicate their thoughts, opinions, or attitudes after utilizing certain items by submitting a variety of reviews utilizing their identities for the specific e-commerce site. As a result, a buyer who want to purchase a brand-new item should first read all of the item reviews [3]. Because of the large number of comments or reviews, it is hard for the client to read and reach a final decision. Furthermore, every consumer must recognize that not each of the reviews given by the customers are real. Many invaders try to create false comments in order to increase or decrease the market rate of a certain type of goods. False reports posted by a person or spammers group have an influence on the sales record of the opponent firm [4]. Several reviews that are completely unrelated to the items are also included in the review database. As a result, before coming to any conclusions, fake or inappropriate reviews must be deleted from the review database. In the last couple of decades, the research group has observed a wide range of technology and science advancements and a growth in internet tasks which including e-commerce, discussion forums, chat rooms, business-making websites, social media, and various other internet activities that could have a positive influence on several research projects towards the development of better decision making support system. Sentimental Analysis (SA) is the analysis of individual's thoughts, opinions, emotions, thoughts, as well as their associative features for evaluating the polarity of the feedback which can be used to improve products and services [5]. Sentimental analysis is also called as opinion mining or sentiment classification is a part of data mining activity which examineswritten content to topics [6]. Sentimental analysis is divided into three categories document level, sentence level, as well as aspect level, which are briefly detailed in the following sections. Sentimental analysis is performed to predict the sentiment within each document, instead of at the phrase level, which doesn't indicate whatever the users like or dislike [7]. A major issue in this domain is text categorization, where a document is termed as an either optimistic or not positive review of a specific entity that might be a film, book, or marketable item. For successful sentiment classification, many machine learning approaches are employed.

Sentiment Analysis is compatible with both the supervised as well as unsupervised learning methods [8], [9].

The technology identifies the sentiment's orientation then retrieves potential elements and sentiment. The aspects are deemed to belong to the similar group, and the algorithm will calculate the ratings for each. The most fundamental job of sentiment analysis is to classify text into distinct polarity groups [10], [11]. Two types of Sentiment classification are binary, especiallywhether the text is optimistic or negative, or multi-class, i.e.whether the text is optimistic, negative, or neutral [12]. Sentiment classification is not the same as text classification. Conventional text classification is done by theme, and there can be numerous groups that are user & application-dependent for a specific document, i.e. there could be only two groups or thousands of groups depending on the app [13]. Text word representations are vital for sentiment analysis, while models such as Bag of Words (BoW) as well as Word Embedding are often utilized [14]. Word embedding is a technique in which words or phrases from a vocabulary are plotted to real-number vectors. The input sentence's words are all encoded as word vectors [15]. Word2Vec is a common example of a word embedding model [16]. The benefit of learning-based techniques is that they are relatively simple & easy to build. For sentiment analysis, customer reviews are significantly more effective than existing model, which fluctuate fast over period.

1.1.1 Types of Sentiment Analysis

Sentiment analysis is classified based on the precise degree of analysis, as represented in Figure 1.1.

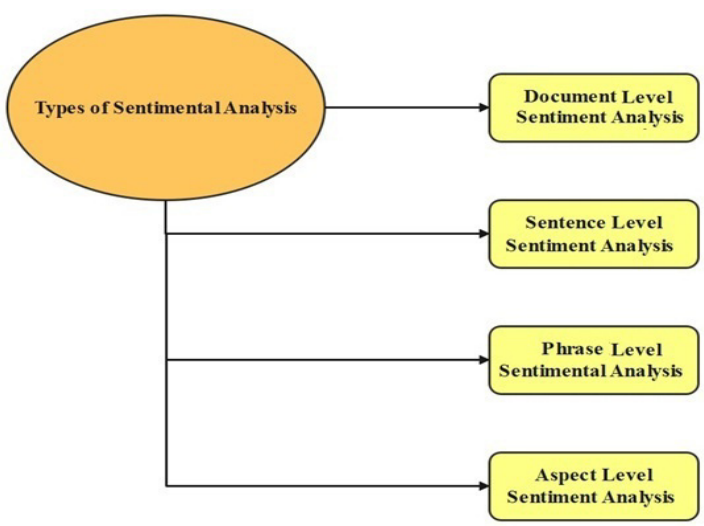

Fig. 1.1 : Types of Sentimental Analysis

1.1.1.1 Document Level

Document level SA is established on the entire file, and the file is assigned as individual antagonism [17]. This pattern of emotion detection is not typically applied. It has the capability to categorize sections or page as good, bad, or unbiased. To categorize the file during this stage, both supervised & unsupervised learning algorithms can be used. The top two critical challenges in document-level SA are cross-domain & cross-language SA. Domain-specific sentiment detection has demonstrated amazing accuracy at the same time domain-sensitive also. The feature vector in such work sare the collection of texts which needs to be domain-specific & restricted.

1.1.1.2 Sentence Level

Sentence level SA determines the sentiment for every sentence separately [17]. It is the method of determining the polarity of a whole phrase without treating every aspect as a separate case and providing an opinion at the sentence level. It is critical in this form of sentiment analysis to determine if the objective phrase is subjective or objective as quickly as feasible, and to determine if a common opinion of the statement is optimistic or pessimistic for objective statements. This type of SA is generally impacted by the sentence's surroundings and is highly important for apps that interact with Twitter posts, Fb posts & replies, brief communications, and so on.

1.1.1.3 Phrase Level

Phrase level Sentiment detection are also conducted, where evaluative terms would be extracted at the phrase stage & categorized. Each sentence may possess multiple or individual attributes. This may be beneficial for item feedback through numerous phrases; thus, an individual component is termed in a statement. In recent times, it has widely emerged as widely explored area of research. Whereas document-level detection concentrated on classifying the whole word as objective, (positive or negative), in this case, sentence-level analysis seems to be multi beneficial due to a file comprises the pair of favorable and unfavorable remarks. The most fundamental basis of language is the phrase, and its antagonism is closely connected to an objectivity of the slogan or words in which it occurs. A statement comprising the descriptive word is usually an objective statement. Furthermore, the phrase selected for expression symbolizes a user's physiological qualities, such as age as well as gender, as well as its motivation, social position, and so on. As a result, phrase is the basis for text sentiment analysis [18].

1.1.1.4 Aspect Level

Aspect Level or Aspect Based Sentiment Analysis (ABSA) is concerned with generating sentiment based on specific features or characteristics of an item depending on feedback. Aspect level is a quite well Sentiment analysis methodology that focuses with establishing the polarities for a specific product aspect. To extract the characteristics (opinion targets) & related opinion words from the provided opinions, the SA process must be built on aspect level. ABSA is comprised of 3 important stages: identification, polarity classification, &consolidation. Aspect identification is an important step in ABSA since it is accompanied by sentiment computation. Factors could be searched directly or by utilizing default implicit features. ML & NLP methods are utilized to retrieve attributes from a word. ABSA is specifically usual in item feedbacks or resort ratings since it aids feedback publishers pinpoint that have numerous crucial components that are significant to individuals and assist them in rectification features that consistsa negativity. This is beneficial to both producers and customers. The collected polarity could differ if they analyse the sentiment values depending on their positiveness or negativeness. Aspect-level sentiment analysis is difficult to extract since it is difficult to determine the specific feature (implicit or explicit) then categorize as per determination.

Therefore, complex methodologies like LSTM, Bi-LSTM, or pre-trained models like BERT and GPT-2 can be employed to accomplish a given task. The experts reject utilizing vanilla RNN because it possesses different problems like vanishing as well as expanding gradient based optimization. Finally, attention-based methods have been employed in factor recognition. During factor diagnosis, the extracted factors' antagonisms are allocated. There are multiple techniques to completing the work. Machine learning algorithms or a human may be utilized. There are numerous approaches to accomplishing tasks. Machine learning approaches or a linguistic-reference method can be applied. Later allocating the antagonism to the feature, a consolidation value could be produced to establish the statement's entire antagonism. The entire sentiment of a phrase is established by either rigid or mild polling. To gauge customer suggestion decisions, customer sentiment is evaluated through qualitative ratings, and consideration of cultural factors [18].

Explicit and Implicit Reviews

Users can post item reviews either openly by identifying the elements & their opinion or implicitly by using metaphors or words regarding the features and their opinion terms. The Aspect Level SA method categorizes reviews depending on their viewpoint orientation, notably Explicit Reviews & Implicit Reviews. Explicit reviews are those in which the reviewer openly mentions the good, bad, or neutral features of the item. Some users could post the review covertly, which is known as an implicit review [19].

1.1.2 Approaches for Sentiment Analysis

Sentiment analysis is carried out by utilizing the procedures indicated in Figure 1.2:

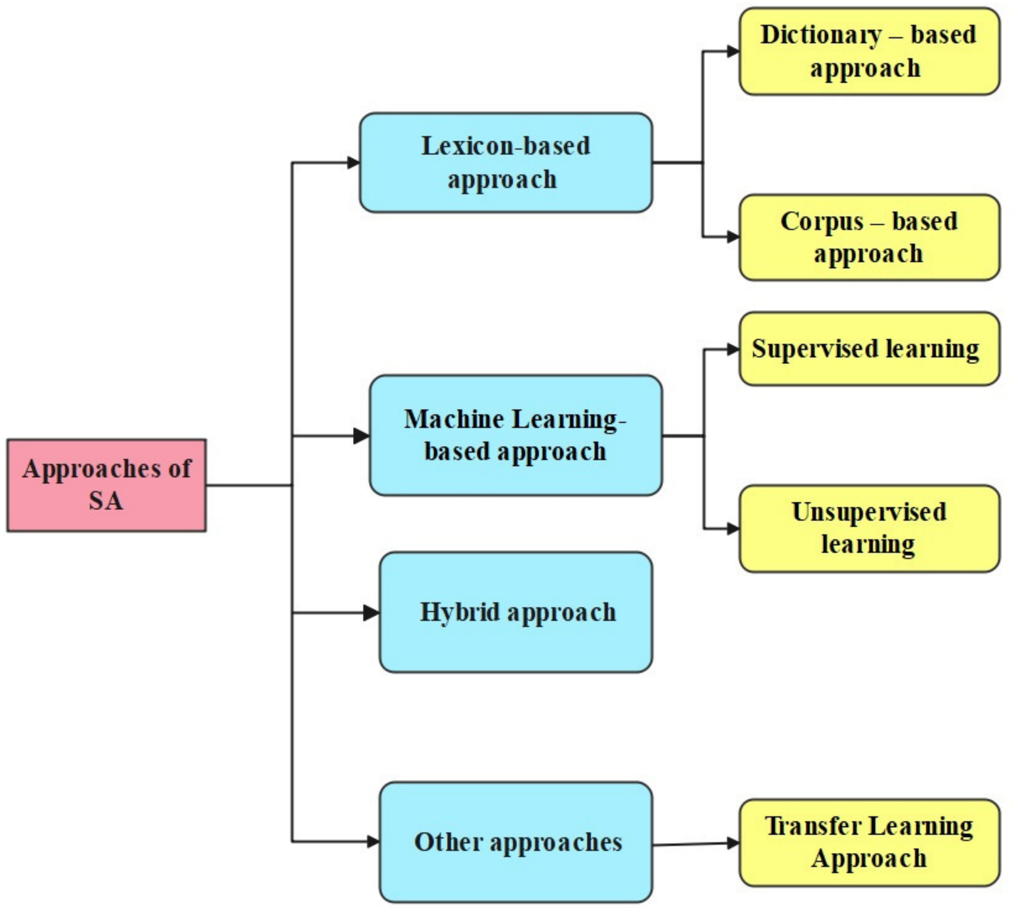

Fig. 1.2 :Sentimental Analysis Approaches

1.1.2.1 Lexicon – Based Approach

Lexicon-based method assesses if a document is favourable or negative depending on the proportion amount of words it includes that are classed as positive or negative. This method of sentiment classification involves comparing text features with a sentiment lexicon that has a pre-defined sentiment value. Those sentient lexicons are frequent opinion words that have previously been defined with a positive or negative value based on their polarity and are recorded in a lexicon. The final value is calculated after incorporating all of the text aspects; if a positive value is acquired from the entire text, the text is categorized as holding positive polarity; if a negative value is acquired from the entire text, the text is defined as containing negative polarity [20]. Lexicon-Based Sentiment Classification has two subcategories. They are as follows:

- Dictionary – based method
- Corpus – based method

Dictionary- Based Method

A lexicon of synonyms & antonyms of phrases is utilized in the dictionary-based technique to generate a collection (count) of lexicons that could be employed for sentiment classification to determine the polarity of an ambiguous word. First, a collection of opinion words with positive & negative polarity is compiled. The collection of words is expanded in the second phase by examining the dictionary for opposites & meanings of the terms. As a result, a final set of lexicons is generated. This vocabulary collection is utilized for sentiment classification, which classifies & determines the polarity of the new document. Word Net, for example, is a lexicon that is employed to create a lexicon SentiWordNet [21].

Corpus-Based Method

Domain-specific dictionaries are created using this method. The method begins with a set of opinions then expands to incorporate other relevant keywords using statistical & semantic techniques. The corpus-based method is divided into two types:

Statistical Based Approach

Statistical approaches are utilized in this technique to find relevant keywords for the seed opinion words. For instance, to find relatively similar terms, Pointwise Mutual Information (PMI) is employed. Whenever a new phrase is met, the polarity of the new phrase is determined by multiplying its PMI value by the collection of seed words [22].

Semantic Based Approach

Similar polarity is given to terms that are conceptually near to one other in this method. The algorithm begins with a collection of opinion words, and the list increases by obtaining opposites & meanings for those seed words. The polarity of a fresh phrase is determined by evaluating the relative amount of favourable and unfavourable terms in the collection of opinion words [22].

1.1.2.2 Machine Learning – Based Approach

In this technique, Sentimental Analysis is performed utilizing Machine Learning algorithm, which entails understanding the processes of classification from existing data defined as training data & then predicting the unlabelled data referred to as test data for sentiment classification. There are two different kinds of ML approaches [23]:

- Supervised learning
- Unsupervised learning

Supervised Learning

This is a directed approach that entails learning from labelled prior data, i.e. the training data contains examples of data with predetermined favourable or negative polarity, and then utilizing this information to forecast the test data for SA. For example, there are supervised learning classification algorithms are Naive Bayes, SVM, DT, RF, and so on [23].

Unsupervised Learning

This is an unintended strategy that entails learning from unlabelled historical data, i.e. the training data has examples that do not contain previously known polarity, but it attempts to learn on its own, i.e. by clustering equivalent occurrences of data termed as clustering and Understanding latent information in unlabelled data and predicting test data for Sentiment classification based on this ability. Unsupervised classification techniques include K-means clustering, PCA, SVD, and so on [23].

1.1.2.3 Hybrid Approach

The mixture of ML and lexicon-related methods for SA is recommended to as hybrid. The hybrid method is frequently employed, with sentiment dictionaries it performing asalientunit in the large number of methods. SA is a hybrid method to polarity detection which involves the pair of quantitative & knowledge-based approach [24].

1.1.2.4 Other Approaches

Transfer Learning Approach

Transfer learning (TL) is an advanced AI method where a previously trained algorithmutilises its wisdom obtained to transmit to a trend method. The comparability of data, division, & activity is utilized in transfer learning. The prior trained aspects are utilized quickly by the trend method that does not necessitate additional training datum. The method may be fine-tuned to a new challenge using training data. This method could be employed to transfer knowledge from one subject to the other. This technique has enhanced in fame as a TL approach because it generates elevated accuracy &results when using much limited time for training than training a recent method from beginning. The transfer learning approach further divided into 3 types which are depicted in Figure 1.3 [18].

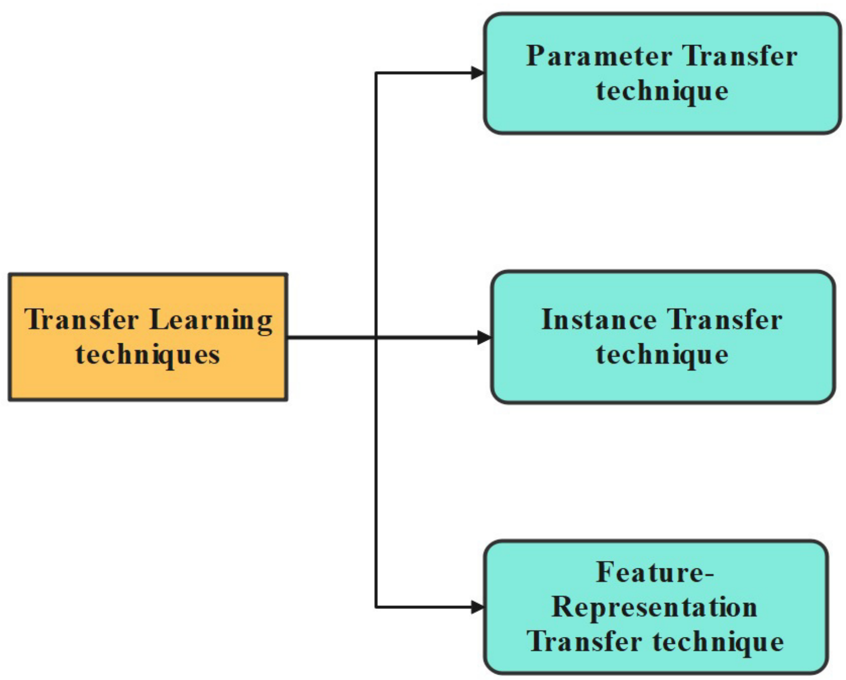

Fig. 1.3 :Types of Transfer Learning Approaches

Parameter-Transfer technique

The parameter-transfer approach makes use of the sources and destination domains' parameter sharing models. Researchers use the parameter transfer approach to transmit trained model parameters from a huge amount of datasets to the objective job. Word2Vec, a basic transfer technique simply for the initial layer of a model, has numerous applications in the parameter-transfer technique deployed to the NLP area, which has a significant influence in reality and is employed in several advanced devices [18].

Instance-Transfer Technique

The data exchange of the source & destination domains is referred to as the instance transfer technique. Re-weighting allows it to be eliminated from the source domain. It allows information from the target domain to be updated by samples from the tagged source domain [18].

Feature-Representation-Transfer technique

The feature-representation-transfer approach requires that the source and destination domains contain some crossover aspects. It is important to feature translate the data from the two sectors into a common feature space before doing typical machine learning. This technique is more extensively utilized & works well in many NLP applications since it has fewer criteria for similarities among the two domains [18].

1.1.2.5 Comparison of Lexicon-Based and Machine Learning-Based Approach

The lexicon-based technique has a benefit over the ML method in that it does not necessitate previous training or adaptable training. It takes lesser time to classify data than a machine learning-based technique. The drawback of the lexicon-based technique over the machine learning method is that it needs management of the dictionary corpus, while the machine learning approach does not. The performance of the ML technique outperforms that of the lexicon-based approach because the ML technique solves the lexical approach's drawback of performance deterioration, it is employed for classification in the suggested method. The accurate allocation of feature extraction, feature selection approach, creating a classifier, and proper interpretation of the input into appropriate class, i.e. in positive or negative polarity, are all challenges for ML techniques.

1.1.3 Importance of Sentiment Analysis

Sentiment Analysis is an effective marketing technique that allows product executives to recognize the thoughts of their customers in their marketing initiatives. It is a key aspect in item or brand identification, consumer loyalty, consumer happiness, the efficacy of marketing and advertising including product adoption. Learning customer psychology may assist managers & customer success managers in making more precise changes to their product portfolio. Emotion-based marketing is a broad word that includes emotional consumer responses including "positive," "negative," "neutral," "uptight," "disgust," "frustration," as well as others. Sentiment analysis is classified into two types. One section is for specific customer service attributes that are based on how consumers utilize the item. This second section covers the item's attributes in the perspective of the experience of customer service. For example, an item could have great reviews in one consumer group but bad reviews in the other, indicating inadequate item or service support.

Product SA necessitates a thorough understanding of how buyers feel. Keywords & terms regularly utilized in customer support calls might disclose information about an item or brand. Utilizing feedback from customer’s tools, a marketing manager or customer success representative could readily explore negative & positive terms utilized in product reviews, indicating concerns with product assistance. Similarly, employing particular keywords connected with product marketing, which including "how to generate more sales" or "earn more money," might disclose marketing tactics. In addition to recognizing & estimating the efficiency of an item roadmap, sentiment analysis may indicate consumer satisfaction if customer service, item use, small problems, & other factors are either strong or weak. Several consumers are unhappy of product assistance after they have bought a product. But, strong consumer service may inspire consumers to buy items from a firm even when the item is suffering unavailability or other troubles. Understanding that sentiments impact consumer decisions that sentiments assist item managers in improving consumer service tactics.

1.1.4 E-Commerce Recommendation Systems

Large e-commerce sites like Amazon.com, eBay.com, etc are still the greatest instances of huge recommender system implementations. Items are often recommended depending on sales, user feedback, and an examination of the consumer's prior purchases [25].

1.1.4.1 Function of E-Commerce Recommender System

RSs have a significant influence on both consumers & e-commerce suppliers. But, the reasons for e-vendors & consumers to adopt RS may differ. One of the primary reason e-vendors implement digital commerce RS, for example, is to improve the quantity of things sold. By delivering highly personalized services, RSs reduce the issue of data overload & assist consumers in choosing a more suitable product. Researchers characterize the motives for service providers to introduce RS to clients as follows:

- Acceptance assists in increasing conversion rates. The basic purpose of establishing an RS in e-commerce is to raise the conversion rate especially to be capable of selling more suggested goods than are typically sold without any form of RS [25].
- Accuracy leads to increased user pleasure. One of the primary responsibilities of RS is accuracy. If a customer thinks the recommended products useful & engaging, the probability of approval is high [25].
- By delivering personalized service, RS enhances customer loyalty, which promotes conversion rates. As a result, regular consumers' choices will be simpler to forecast, increasing the accuracy of recommended goods. Furthermore, in order to acquire new users' loyalty, the system must be able to detect new customers & properly deliver services [25].
- RS assist in the sale of more different commodities. One of the reasons that the user appreciates using the RS is that it recommends various related goods. As a result, variation boosts customer loyalty and the amount of various products being sold [25].

1.1.4.2 Recommender System Algorithms

To analyse the advancement of E-Commerce Recommendation System, we should first evaluate the primary state recommendation approaches, which including Content Based, Collaborative Filtering-based, & hybrid methods, as shown in Figure 1.4.

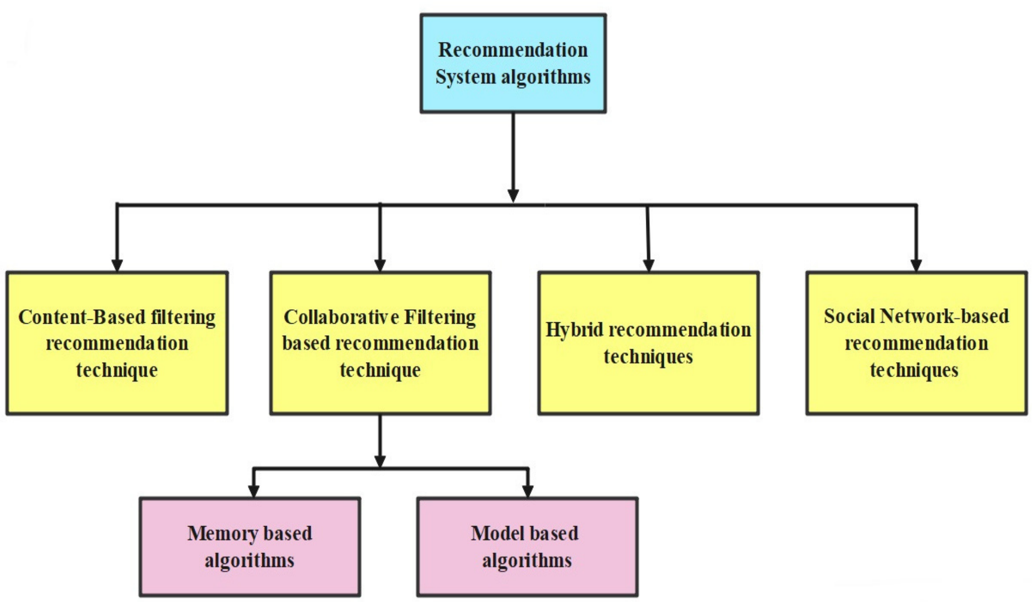

Fig. 1.4 : Recommendation System Algorithms

Content-Based filtering recommendation technique

The Content-Based (CB) filtering suggestion approach suggests things relevant to those bought previously by the consumer. The CB recommender approaches' difficulty is analysing the quality of the products liked by a certain customer to establish the customer's choices. Two strategies are often employed to create suggestions [25].

- Employing conventional data retrieval approaches such as cosine similarities, TF-IDF, and LDA.
- Utilizing modern ML techniques like Nave Bayes, vector support machines, and decision trees.

There are just a few restrictions to content-based filtering recommendations: Overspecialization occurs whenever a recommender is unable to suggest unexpected, yet acceptable things; soliciting consumer feedback occurs whenever the recommendation performance can only be enhanced with additional historical data from the customer. For instance, if a consumer purchases, reviews, or writes comments on an item [25].

Collaborative Filtering based Recommendation Technique

Recommender systems depending on Collaborative Filtering (CF) propose a product for a specific consumer depending on things formerly liked by the other consumers. CF techniques are classified into two types:

- **Memory-Based Algorithms** – The recommendations are depending on the opinions of the closest neighbours. A consumer gets suggestions that are comparable to those he evaluated before in the item-related CF technique. In the user-related CF method, suggested goods are depending on comparable customers. Pearson correlation coefficient and cosine resemblance metrics may be utilized to determine the similarities among customers or things. Empirical evidence shows that the item-based CF method surpasses the user-related method [25].
- **Model-based Algorithms**- SVD, tensor factorization, and Bayesian Networks (BN), produce recommendations by examining hidden characteristics in customer reviews or by developing a model to anticipate the most desirable product that a consumer may like to buy. CF-based approaches are commonly employed in the development of e-commerce RSs. On the other hand, CF faces issues which

including cold start, scalability, & data sparsity. Data sparsity is defined as a low percentage of rated things relative to the total amount of products. A cold start issue arises when there are fresh customers or items with limited history activity [25].

Hybrid Recommendation Techniques

In order to minimize restrictions including the cold-start problem, hybrid recommendation approaches incorporate two or even more recommendation techniques, which including CB & CF-based methods. Weighted, switching, mixed, cascade, etc are the methods for creating a new hybrid RS. Statistical approaches which including NN, BN, clustering, hidden features are commonly used in hybrid filtering. Although hybrid methods provide an alternative for several conventional RS methods, they need more knowledge and work to deploy [25].

Social Network-based Recommendation Techniques

Social network-based (SN) recommender approaches employ data from the social networks, such as consumer preferences or social buddy connections, to increase recommendation accuracy & solve significant difficulties, such as cold-start & data sparsity issues. This strategy is dependent on the idea that social partners have similar interests. By analysing customer ratings, SN generates community-based suggestions. SN generates community-based recommendations by comparing ratings and reviews. Bayesian networks & neural networks are examples of probabilistic approaches used in SN recommender systems [25].

1.2 Background

Facts are certainly crucial in real life for people or companies to reach decisions or make judgments. While facts are important, attitudes and opinions are as important. Human conduct is heavily influenced by their unique emotional state & opinions, which including approach, mood, or sentiment. Not just human behavior, but also the decisions we make, may be influenced by the observations and beliefs of others as expressed in the form of their opinions. Even though not all persons are expected to become arrogant equally by ideas, opinions, & actions, perception is among the fundamental features of all social creatures. Because of financial & social considerations, some companies and people are thrilled by the feelings of others. Others' feelings can easily manipulate people's opinions about their belongings. As a result, "what the other people may think" has become an important factor in decision-making. As a result, it was critical to look deeply into people's feelings about a certain topic of interest. People now have the ability to convey themselves digitally as internet usage has increased. The research to do sentiment analysis have become much more blooming as the number of opinions expressed internet has increased.

Prior internet opinions, people utilized to communicate their thoughts by word-of-mouth. People frequently speak more about things they purchase from friends or family. They also consider these evaluations while making their individual purchases or recommending something to another person. The development of websites that focus user-generated material allows individuals all over the world to submit their opinions or views on a variety of topics. The internet technology has transformed word-of-mouth recommendations into broadcast communication. Online opinions in the form of reviews, comments, & ratings provide a chance to learn about and make decisions based on the attitudes and emotions of individuals who are not linked to one another. It is now quite simple to exchange emotions and observations on the internet. It has become quite simple to communicate sentiments on the online and watch other people's ideas, resulting in the development of an increasing amount of data comprising opinions in multiple kinds which including online reviews, blog entries, forum conversations, as well as various social networks. Any data containing unfavourable or positive claims about the real product or the production plant might have an impact on the firm's profit or sale margins.

1.3 Challenges

Sentiment Analysis presents a number of difficulties, ranging from processing expense to unstructured writing as well as the prevalence of lexical variants. Figure 1.5 depicts the challenges of SA.

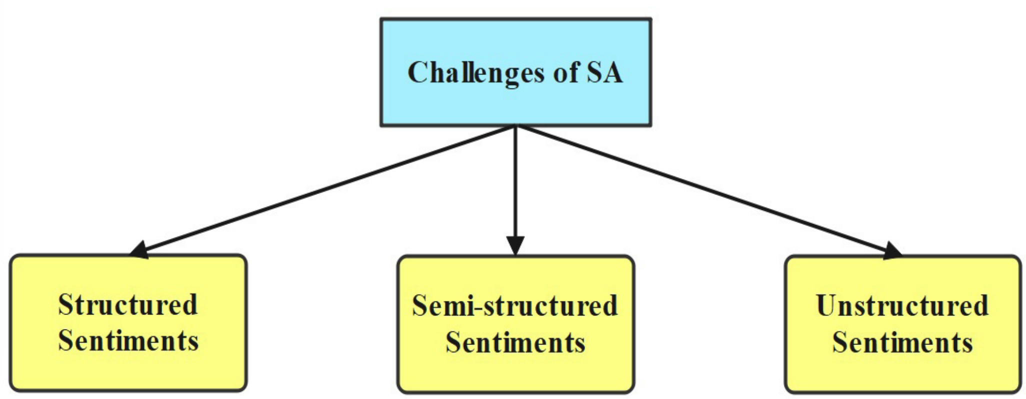

Fig. 1.5 :Challenges of Sentimental Analysis

1.3.1 Structured Sentiments

Formal sentiment review provides structured feelings that they are mainly concentrated on formal concerns which including books, research, etc. Since the writers are specialists, they provide ideas or views about scientific or actual issues [18].

1.3.2 Semi Structured Sentiments

Semi-Structured Sentiments are intermediate sentiments among structured & unstructured sentiments. These need an understanding of several feedback-based problems. The writers describe the merits and shortcomings of this approach individually, and the pros & cons portions are often consisting of huge phrases [18].

1.3.3 Unstructured Sentiment

Unstructured Sentiment is one of the unrestricted& casual composition way where the author is not bound by each restrictions. The text may have numerous phrases, every one of which may contain both merits and drawbacks. For instance, Unstructured reviews provide additional opinion data than their formal equivalents. A property that is stated clearly: If a feature appears in the segment/chunk of a review phrase, it is considered to be an item's explicit feature. For instance, the visual in the part is fantastic. The picture is a prominent feature. An implicit feature of an item is one that is not expressly listed in the review portion but is assumed [18].

1.4 Applications

Sentiment analysis has several uses, including analysing consumer opinions and customer psychological health depending on social media comments. Meanwhile, technology advancements such as Big Data, IoT, & Cloud Computing have expanded the scope of Sentiment Analysis applications, enabling it to be applied in nearly any field. Figure 1.6 depicts a few of the most often utilised applications in sentiment analysis. The following are some of the most important fields or sectors in which Sentiment Analysis is used [18]:

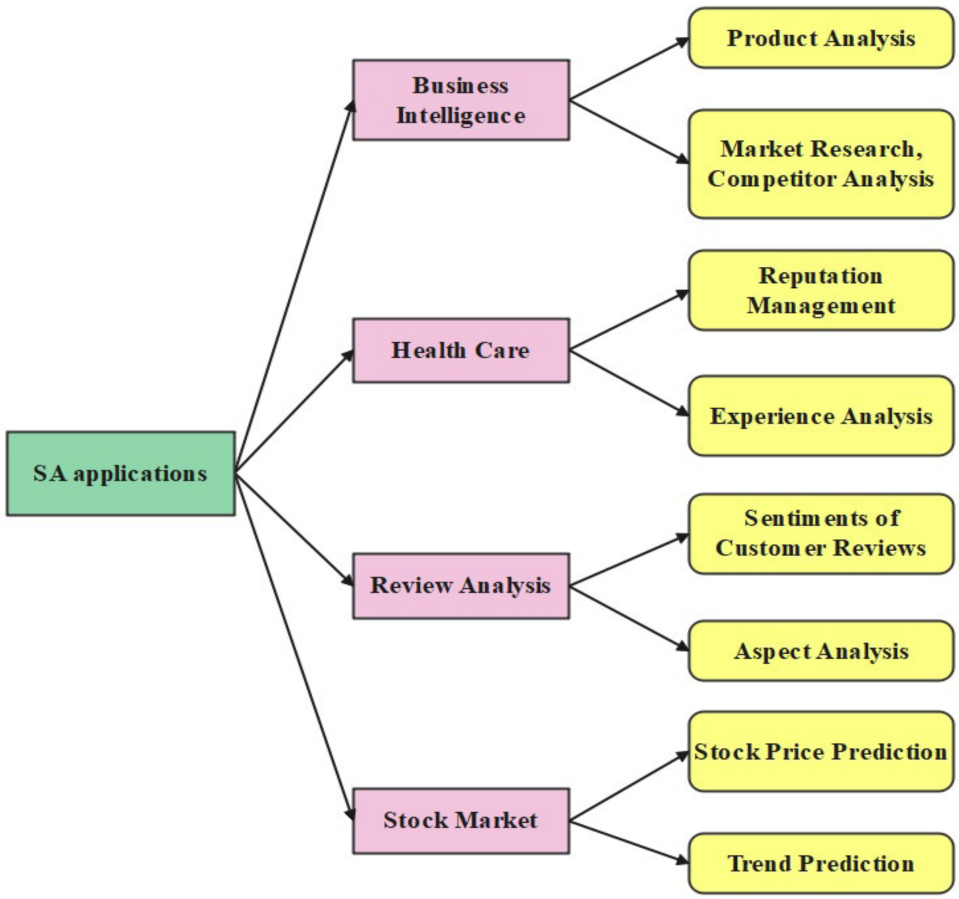

Fig. 1.6 :Applications of Sentimental Analysis

1.4.1 Business Analysis

Sentiment analysis in business analytics has various advantages. Firms may also use sentiment analysis information to enhance items, investigate customer comments, and design a new marketing approach. A most common application of sentiment analysis in business intelligence is to evaluate client opinion of goods or services. But, these research aren't restricted to manufacturers; customers can utilise them to analyse things and make better educated purchases. Customers can also utilise them to compare products and make better informed decisions. For the past eight years, Researchers have been providing meal evaluations on Amazon.com. Where. Emotion lexicon divides them into 8 emotions & 2 states (positive & negative). In order to detect consumer behaviours and risks, as well as to improve customer satisfactions, researchers discovered that the SA could be utilized [18].

1.4.1.1 Product Reviews

The quantity of items supplied and consumer reviews are increasing as the e-commerce industry grows. One of these is sentiment analysis, which can assist consumers to select better items. Product reviews are subjected to word or aspect level sentiment analysis. Sentiment analysis could be used to assess what customers consider about a company's current product after it has been launched, as well as to examine reviews and comments. Keywords for a certain product aspect could be selected, as well as a sentiment analytical method could be trained to recognise& assess the required data [18].

1.4.1.2 Market Survey and Opponent Analysis

Market survey, along with brand image monitoring & customer opinion analysis, is likely a most popular sentiment analysis application. The goal of sentiment analysis is to evaluate who is rising in popularity amongst opponents as well as how marketing initiatives differ. It might be used to create an in-depth picture of a brand's as well as its opponents' customer bases from the bottom up. Sentiment analysis might gather information from many sites like Twitter, Facebook, & blogs, give practical results, and effectively handle business intelligence challenges [18].

1.4.2 Healthcare and Medical Domain

The medical domain is the most popular area has recently used sentiment analysis. Information are gathered from a variety of sources, including questionnaires, Twitter, blogs, news stories, reviews, and so on. This data is then evaluated for a variety of purposes, including regular analysis & investigation of latest clinical updates. Domain specialists are continually investigating new applications for sentiment analysis as well as other NLP applications. This application assists healthcare providers in collecting and analysing patient moods, epidemics, adverse medication responses, or disorders in order to enhance health services. Some authors highlighted the difficulty in employing sentiment analysis in health care due to the domain's particular and distinct terminology. Twitter posts about patient experiences can be used to supplement global health analysis. With Twitter's Streaming API, they created over 5 million breast cancer-related comments over the course of a year. The comments were categorised using a typical LR classifier as well as a CNN model following pre-processing. Positive treatment experiences, gaining public support, & raising public knowledge were all associated. [18].

1.4.2.1 Reputation Management

Brand monitoring & reputation control are two applications of sentiment analysis in various industries. Fashion firms, marketing agencies, IT firms, hotels, media, as well as other organisations benefit from assessing how users evaluate their brand, item, or service. The sentiment analysis tool offers additional diversity & intelligence to the depiction of the brand and its items. It allows companies to monitor how their consumers evaluate their brands & reveal specific facts about their opinions. Identify trends or developments, and pay more attention to the presentations of experts. Overall, sentiment analysis is also used to automate both the media surveillance network and the security system that corresponds with it. Maintain an eye on the brand's conversations & ratings across variety of social media channels is significant [18].

1.4.3 Review Analysis

Sentiment analysis is widely employed within the entertainment realm whereas film, programme, or short film comments can be evaluated to assess the audience's reaction. This not only allows audiences to make smarter decisions, and also it helps decent material obtain popularity. Sentence level SA is often utilised in this area to properly estimate the total sentiment of reviews submitted. The tourism industry has attempted to increase customer's experience by establishing machine learning as well as online customer recommendation systems depending on smart, data-driven decision-making methods. It has also been discussed whether human activities should be categorized as either optimistic or negative reliant on online feedbacks posted by useful consumers [18].

1.4.4 Stock Market

Stock price prediction is an example of the areas of sentiment analysis. It is able to do so by examining current stock market data & forecasting stock price movements. Data may be gathered from a variety of sources, including Twitter, news articles, blogs, and so on. Sentence level sentiment analysis may be performed on these words, later that the total polarity of letters of news from a certain firm will be determined. Good news tends to drive an upward trend, while negative news drove a reduction. Bitcoin as well as alternatevirtual coins are linked to Block chain, a unique technology. Members in the block chain network use peer-to-peer voting procedures to validate digital transactions. However, studies that use SA to the field of block chain technology were always rare [18].

1.4.5 Customer Feedback

Integrate and evaluate every user input from call centres, emails, surveys, chats, as well as the web. Sentiment analysis will enable data classification and organisationfor the purpose of uncoverrecents including recurring problems and concerns. Sentiment research might help find a specific client group and then construct a value offer, both of which are vital aspects of a profitable operation [18].

1.4.6 Social Media Monitoring

Sentiment analysis of social media information analyses customer sentiment 24/7 a day for 7 days per week, in actual-timewhilewhatevernegative begins to propogate. Thus allowing for a quick response or image strengthening when positive reviews are received. This also provides constant, trustworthy information on clients, allowing the decision-making process to follow growth from season to season [18].

1.5 Objectives

- To obtain a precise recommendation of products by assessing the product reviews posted by the customers, a new opinion analysis framework is created.
- In order to enhance classification execution, the Word Embedding Attention (WEA) strategy is introduced in the Bi-directional Long Short Term Memory (Bi-LSTM) method.

1.6 Problem Statement

- Because of the fast rise of electronic commerce, research on online evaluations has become a well-established area. However, unstructured big data through online reviews has both harmful and beneficial effects on customers.
- The fusion of features creates overfitting technique and irrelevant features that impacts the classification execution of the suggested method.
- Before purchasing a product, more number of users frequently checks the online reviews. As a result, early reviews on before purchase have made a huge impact on product sales.

1.7 Motivation

In recent years, the number of User-Generated Content (UGC) has increased considerably because of the simple accessibility of internet. This UGC has the greatest impact on the e-commerce industry since consumer opinion, feelings, product ratings, or comments in the type of online user reviews all have an influence on the development of any firm's operations. They also assist customers in deciding whether or not they will purchase an item. However, if a buyer checks only a few comments out of thousands, he or she could acquire biased and fail to make the appropriate selection. As a result, consumer sentiments, which constitute a significant piece of data, should be appropriately analysed& handled. However, manually analysing those thousands is laborious. However, physically analysing large numbers of reviews is challenging. For this reason, sentiment analysis is being used, which is an analytical field of research that systematically understands the sentiment, feelings of the consumer from internet user generated content as well as classifies/predicts the consumer reviews as positive, negative, or neutral. The study's objective is to investigate different sentiment classification learning algorithms in order to develop more efficient methods to utilize this customer data.

1.8 Description of the tool

Python is a general-purpose, high-level scripting as well as programming language. Its model concept emphasis code readability through using extensive indentation. Python is garbage-gathered& dynamically typed. It allows for a vast variety of programming techniques, involving procedural, object-oriented, as well as functional programming. Due the extensive standard library, it is simply defined as a "batteries included" language. Guido van Rossuminitiateenhancing Python as analternate towards the ABC computer language in the 1980s, and it was really launched in 1991 as Python 0.9.0. Python 2.0, which was launched in 2000, appendedtrend capabilities which including list comprehensions, cycle-detecting garbage collection, reference counting, as well as Unicode support. Python 3.0, established in 2008, represented amajor update that did not maintain backward-compatibility with former versions. Python 2.7.18, published in 2020, was the final Python 2 version. Nowadays, Python is placed as one the top programming languages.

1.9 Organization of the Thesis

The emphasis of the thesis's focus on classification of customer review using sentimental analysis in the E-commerce websites. This thesis is presented with the following general arrangement, In Chapter 1, the overview of Sentimental Analysis is provided. It provides a concise overview of the types of SA, various method used for SA, applications, challenges, motivations, background and tool description.

In chapter 2, An overview of the literature on the ideas and difficulties of Sentimental Analysis is presented. A thorough literature review of various methodologies also presented.

In chapter 3, A new opinion analysis framework is created for precise recommendation of products by assessing the product reviews posted by the customers. By performing quantitative as well as comparative analysis, the proposed model produces a well-defined results This section also offers an extensive summary of the experimental results.

In chapter 4, WEA-BiLSTM model is applied on the Amazon dataset for the sentiment analysis. The WEA-BiLSTM model has higher efficiency for less number of training data due to model uses CNN feature extraction for classification. This section also offers an extensive summary of the simulation's results.

In chapter 5, it provides a short overview about the complete work and its future progress. Contributions must be enlarged, as well as achievements must be calculated at the end.

CHAPTER – II

LITERATURE SURVEY

The World Wide Web (WWW) creates large amounts of data in the structure of user views, feelings, decisions, and discussions regarding various social events, goods, brands, and politics through the use of social networks, newsgroups, review sites, and blogs. Users' online sentiments have a significant impact on readers, product suppliers, and politicians. It is necessary to assess and well-structure the unorganized form of social media data, and sentiment analysis has drawn a lot of attention for this reason.Text organising technique known as sentiment analysis is used to categorise expressed attitudes or emotions into distinct groups, such as positive, negative, favourable, unfavorable, thumbs up, down votes, especially the challenge for sentiment classification is shortage of sufficient labelled datum in the domain of Natural Language Processing (NLP). And also because deep learning methods are effective because they have the potential to automatically learn, sentiment classification and deep learning approaches have been combined to address this problem.

2.1 Product Opinion Analysis for Websites using Deep Learning

Li Yang and Ying Li[26] Convolutional Neural Network (CNN) and attention-related Bidirectional Gated Recurrent Units are combined in the newly created and proposed sentiment analysis model. This layer's primary job is to take the input matrix's context characteristics and extract them. Often used to analyse sequence information, the GRU method is a variation of the recurrent neural network method. The present output may be influenced by combining historical data from earlier moments, and it can also extract context elements from sequence data. We utilize the BiGRUmethod to retrieve the contextual aspects of the given text since both the words before and after the current word in the text data impact it. Then use CNN to retrieve the key features from the given matrix, followed by BiGRU to take into account the text's order detail and extract word context features, learning algorithm to assign varying weights to the various input features, highlighting text sentiment features, and connected directly to classify sentiment features.The benefits of deep learning method and sentiment lexicon outweigh the drawbacks of the current sentiment analysis approach for product reviews. The deep learning-related technique, however, does not call for any kind of physical involvement. Through neural network structure, it can spontaneously choose and extract features and learn from its mistakes.

Lin Li and Tiong-ThyeGoh[27] presented a study that looked at a neglected yet important research subject on sentiment analysis in internet reviews. Word embeddings and a linear learning algorithm were employed in their suggested deep learning-based classifier. In order to combine data from various sources, they also suggested ensemble of two approaches and two methods that include the pair of visible and deep characteristics. Finally, they empirically shown that mixing data from several characteristics and analyzers may significantly boost sentiment categorization performance.To improve the precision of the sentiment categorization on the reviews, a huge dataset of movie reviews is processed using Long Short-Term Memory (LSTM) with gating techniques. To regulate the inner world and outputs at every time, the LSTM network was created alongside of three rectified linear units: the inputs gate, output stage, as well as output gate.As a result of the LSTM model's significantly higher execution when distinguished to the SRN and CNN, it is able to handle long-term dependencies and remember long sequences of information better. This is accomplished through the communication of the three gated units, which determines which information is output during in the current state, is retained, and is discarded. The LSTM is more advantageous in many situations since it is somewhat insensitive to the gap length. Long, however. The variations in the classifier with regard to changing the sentence length, however, could not be summed from their experimental data.

Mohammad EhsanBasiri and MoloudAbdar[28] found a brand-new way for classifying medicine evaluations' sentiment utilising an association of deep learning and machine learning methods. To assess the medication reviews, two deep fusion methods related on three-way decision theory are proposed in this paper. The first 3-way fusion model between a deep learning algorithm and a traditional learning algorithm (3WDT) was created utilizing deep learning as a main categorizer and traditional learning as a subsidiary approach that is utilized when the deep learning algorithm's confidence is low during the classification of test samples. Three deep and one traditional model are trained on the full training set in the second suggested deep fusion method, called 3-Way fusion of three Deep models with a Traditional model (3W3DT), and each classifies the test sample independently. Results were greatly enhanced by the 3W3DT model. When compared to other approaches to SA in medication reviews, the prior stages' usage of NB and learning techniques algorithms produced better results. The suggested models can be used to solve additional classification issues of a similar kind, like sentiment identification and rating predictions in the SA field. Nevertheless, the confidence computation mechanism should be used for the particular situation in order to apply the suggested solution to such issues. Additionally, it is important to carefully evaluate the sample space limits in order to split it into the necessary number of classes.

Wei Li and Luyao Zhu[29] put into practise a model that we proposed: CNN-LSTM and CNN-Bidirectional LSTM, two-channel Convolutional Neural Networks with Long Short Term Memory. In order to create the given data specimen of a constant size and increase the quantity of sentiment deytail in individual feedback, first suggest sentiment padding, an unique padding strategy in comparison to zero padding. The sentiment buffer and parallel two-channel CNN-BiLSTM structure make up the model's two main contributions. Sentence padding was created specifically for text SA tasks. The gradient vanishing issue between the input nodes and the initial hidden layer is less problematic with sentiment padding than with zero padding. Additionally, unlike with a grayscale map, zero padding in vector space does not mean "none." Sentiment padding increases the amount of sentiment data in a review, which is helpful for classifying sentiment polarity. While the BiLSTM model can process lengthy sequences, the CNN model excels at extracting local features. In this situation, combining CNN and BiLSTM simultaneously will yield useful information for both systems.Incomparison to CNN-BiLSTM model, CNN-LSTM model performed better on a dataset of reviews on Chinese tourism. On the Chinese tourist dataset, the CNN-BiLSTM model was unable to even outperform its single branch BiLSTM. This shows that the execution of the parallel two-channel method relient on the coupling among the two branches and unrequired on the execution of its individual section.

Somya Ranjan Sahoo and B.B. Gupta[30] foundanapproach for automatically detecting bogus news in the Chrome environment that can do so on Facebook. To identify bogus news, our suggested method employs machine learning and deep learning-based analysis, together with numerous variables connected to the user's profile. Our suggested method detects bogus news in a proactive manner. Additionally, our method gathers numerous news articles shared by Facebook users and other attributes connected to various profiles to analyse bogus activity.Thefundamental definition of deep learning is that it is a type of machine learning which learns to depict the world as a layered hierarchy of ideas, with individual idea connected to the previous one with abstract model. Deep learning classifiers utilise their own data analysis for learning. Our method for predicting decisions benefits from the selection of characteristics based on user profiles and shared content. Our crawler and the Facebook API extract a number of new features. The fundamental definition of deep learning is that it is a type of machine learning that learns to describe the world as a layered hierarchy of concepts, with each concept connected to the earlier one with abstract concept. Deep learning classifiers utilise their own data analysis for learning. For enhanced decision-making, the gathered data is analysedutilizing a number of machine training and deep learning-based tools.The most recent information, however, is dubious and frequently misleads other users on a social media site. It is demanding for detection method to find fake news related on shared resources since disinformation is purposefully distributed to induce viewers to accept false news.

Erick Kauffmanna and JesúsPeralb[31] In order to aid marketing managers and customers in the decision-making process, a modular framework built on sentiment analysis and the crucial problem of false review identification was developed. In order to obtain sentiment values, a novel variable that sheds insight on consumer behaviour, the framework analyses consumer evaluations using NLP technology and gives extra and comparative information. On assessments of goods from high-tech businesses, the FRDF was put to the test. According to consumer sentiment, brands were graded. The results show that when combined with the star score, marketing teams and consumers will find this tool valuable. Tools for detecting false reviews and sentiment analysis are the foundation of the FRDF. In essence, more details are taken from user evaluations and used to adjust the initial star rating, if necessary. Marketing managers, resellers, and customers' decision-making may all be informed and hence improved by FRDF. Big data product reviews and the pertinent details of each product, such as the price, the brand, and the product categories, are fed into this framework. This information is examined to produce fresh market knowledge that will aid in the decision-making process for managers and customers alike.The framework offers additional and comparative data extracted from customer reviews and methods them using NLP technology to get emotion values, a new variable that illuminates consumer behaviour. We have thoroughly reviewed the prior work on the topics related to this framework, including big data and advertising, sentiment analysis, and fake reviews. The results are outlined in the background section, where our efforts to the state-of-the-art are also included.

Md Rafqul Islam and Shaowu Liu[32] based on multiple DL approaches, give a state-of-the-art comprehensive evaluation of the current issues, fixes, and verification of Misinformation Detection (MID) in digital social platforms. The goal of this study is to present the state-of-the-art DL as an evolving approach on huge social network data, as well as to highlight current and future trends in MID research. In order to best grasp the situations of a current issues well astrendanalysis concerns, a systematic review is therefore essential for successful of DL for MID the pair in academia and industry. The research on MID has been reviewed and summarised in some very excellent ways, however there are still enough gaps for a more thorough assessment of the misleading literature. To comprehend the rationale for using DL approaches on MID and jointly advance the state-of-the-art, datasets are required. Several well-known strategies have been employed for MID in various fields, however they are not interchangeable. The data gathering may differ greatly due to different study goalsMID, where DL is used to analyse data autonomously, build patterns, and make decisions that improve outcomes in addition to extracting global characteristics. In addition, it is challenging to collect the data from the existing study work that is available. For example, some datasets primarily address personal problems, at the same time alters include administrative, commercial, and socially significant problems.

Anwar Ur Rehman and Ahmad Kamran Malik[33] To address the outcomes of the sentiment analysis issue, a hybrid model called the Hybrid CNN-LSTM Model was built employing LSTM and a quite deep CNN method. Utilizing convolutional layers as well as max-pooling layers, the CNN method effectively recovers higher level information. The polling layer in the CNN model is used to lessen the computational complexity. CNN uses polling algorithms that minimise output size from individual stack layer to the ensuing while preserving crucial information.Although there are other polling methods obtainable, max-polling is most frequently employed when the pooling window has the max value element. The long-term relationships between word sequences are capable of being captured by the LSTM model. The suggestedhybrid CNN-LSTM method's primary architecture uses a corpus as its input and, during the pre-processing stage, executes tasks including statement separation, tokenization, stop word removal, and stemming. Following that, Word2Vec's word embedding layer is applied. The LSTM layer recognises long-term relationships amongtexts, whereas the convolutional layer extracts high level characteristics. In the end, we use the sigmoid function to apply a classification layer.In relation to the accuracy, the CNN-LSTM method outperformed the standalone CNN and LSTM models on two benchmark datasets of film feedback. The results of the empirical demonstrated that the suggested strategy improved categorization in terms of accuracy and decreased error rate. However, the issue with this technology was that deep learning algorithms cannot directly interpret human text.

W.M. Wang and J.W. Wang[34] developed a heuristic deep learning algorithm that classifies emotional judgments into seven pairs by extracting them from consumer product evaluations. using the heuristic deep learning approach to extract and categorise various emotive characteristics from online user product evaluations. Online reviews are gathered and examined for emotional design. They employ syntactic and semantic text mining techniques, which require a specified lexicon for mining and extraction, to automatically extract product attributes and their accompanying emotional responses from online product descriptions and customer evaluations. Their studies concentrate on the extraction of emotive words and phrases from customer evaluations, comparisons of the viewpoints of designers and customers, and summarizing of emotive thoughts.A deep learning approach to sentiment analysis. Numerous studies already in existence categories reviews as favourable or unfavourable, score the reviews, or summarise viewpoints, which is inadequate for affective design. In order to build the training set and additional measurements for rule formulation and model training, it first gathers online customer reviews, diagram above illustrates the feedback, and extracts important information. Second, it creates text analysis rules and classifier using the training data. Third, it evaluates how the suggested strategy, which mixes rules and models, performs in comparison to other approaches now in use. In comparison to the rule-based extraction approach and individual machine learning models, the combination of rule-based extract and deep learning models performs better. It demonstrates how the heuristic combination technique works to better utilise the benefits of regulation extract and machine learning models. Under sampling might harm overall performance, while oversampling could cause an overfitting issue because of the small amount of the training data.

Satyendra Kumar Sharma and Swapnajit Chakrabor[35] created a forecasting study for book sales on the Indian Amazon marketplace. This type of prediction skill, Amazon, is also crucial to effectively managing the supply chain and ensuring consumer happiness. Regression analysis, decision-tree analysis, and artificial neural networks are three modelling approaches that are tested in this study to see how well they forecast book sales on Amazon.com by taking into account a variety of pertinent variables and their interactions. Online reviews that are included as predictors in these models undergo sentiment analysis to determine their polarity. On Amazon, the product reviews are categorised and shown page by page. Both the review title and the review content are taken from the first two pages of each review for sentiment analysis. Additionally, the regression analysis produces equivalent findings both when sentiment and interaction components are included and when they are not. The comparison of these models yields numerous important conclusions. First off, all three models support the idea that review volume is the single most significant and crucial predictor of book sales on Amazon.in. Second, the influence of discount rate, discount amount, and average ratings on sales forecast is negligible to considerable.Thirdly, whereas both positive and negative review sentiment are individually significantly predictive according to regression and decision-tree models, they are completely unsignificant according to the neural network model. The existing study suggests that both positive and negative sentiment are relevant, with the former having a greater effect in forecasting sales, which is at odds with the neural network method's observation. However, to overcome the issue of the broad range of values, sales Rank is converted using natural logarithm.

Susan A.M. Vermeer and Theo Araujo[36] found a propose and test a strategy based on supervised machine learning that determines if electronic term eWOM is relevant for the brand to respond before classifying the seven distinct categories of eWOM. This study determines eWOM that is deserving of a response based on its substance rather than its attitude. The findings show that when compared to any type of sentiment analysis, these machine learning algorithms discover significant eWOM on social networks with a great deal more accuracy. categorised the many eWOM kinds into three primary groups.neutral, satisfied, and dissatisfied. First, the category of Dissatisfaction includes consumer complaints or rejections of the product, service, or brand as a whole. Second, acknowledgements and/or compliments are included in the Satisfaction category. Finally, everything in-between the two was labelled as Neutral, including remarks, inquiries, and/or ideas. Take management concerns into account when evaluating the outcome for the purpose of making the most of the characteristics that machine learning approaches have to offer and, more significantly, in order to determine which of the algorithms has done better. However, dictionaries find it challenging to evident following from word co-occurrences, exactly as sentiment analysis. Dictionary-based text analysis frequently runs the danger of being overly particular and omitting terms since the method depends on a number of subjective processes; as a result, it could not accurately reflect the complete data set or be as adaptable to be applied to new texts.

Monika Arora and Vineet Kansal[37] Conv-char-Emb is a deep convolutional character level embedding neural network model that was built as a text normalizing technique for SA of unstructured data. The suggested technique accomplishes both sentiment normalization and classification for unstructured texts. CNN is frequently utilized for sentiment categorization, which does not need to be familiar with the semantics or syntactic constructions of a certain language. The ability of the software to handle messages in several languages is enhanced by this. The character level embedding approach is used to train CNN's deep convolutional architecture to evaluate the sentiment of the phrase. In order to examine the sentiment of the sentences, the suggested technique determines the polarity of the sentence without relying on a lookup table or word2vec task. Preprocessing is a crucial step for precise polarity recognition because the input for our sentiment analysis work is noisy data from Twitter. Thus, the unstructured input is standardized before word level CNN adaptive learning analyses the emotion. The RNN network ascertains the texts' long-term relationships, whereas the CNN model learns their local properties. The key component of the phrase was then extracted using a pair-wise prepared by this method that computed the similarity across the text. CNN outperforms other models in terms of accuracy measure gain thanks to normalization and learning-based sentiment analysis. CNN, however, needs specific knowledge in order to train its network to estimate attitudes.

Santosh Kumar Banbhrani and Bo Xu[38] For the purpose of predicting review ratings, a Spider Taylor-ChOA: optimised Deep Learning based sentiment categorization was created. RMDL based on Spider Taylor-ChOA is used to categorise the emotion. Taylor-ChOA spider The SMO and Taylor-ChOA are combined to create the planned spider Taylor-ChOA. The SMO in this case takes its cues from spider monkeys. It is dependable among swarm intelligence techniques and provides self-organization that simulates a reaction at the global level through interactions between small units. It improves swarm intelligence that is incorporated.The SMO has the capacity to identify improved solutions and the global optimum. The determination of the user's sentiment intensity toward the target products from various forms of reviews is the goal of the necessary sentiment analysis method known as review rating prediction. The objective is to develop a method for predicting review ratings that is based on sentiment categorization.By managing the convergence speed, it enables the ability to strike a balance between exploitation and exploration. The SMO is hence quite sensitive to hyperparameters. The Taylor-ChOA approach was created by integrating the advantages of both ChOA and the Taylor notion. It covers both confined and unconstrained situations and aids in resolving the issue of convergence speed. However, the method was not tested while taking into account different classes and review infrastructures.

Jie Chen and Jingying[39] proposed an improved hierarchical neural net for document-level sentiment classification on the user's review behaviours. Both the training set and the test set's document reviews are broken down into user groups, and all of each user's reviews from the training set are utilised to train a Long short - term memory hierarchical neural network to produce feedback content representations that take the user's review behaviours into account. The resemblance of the review habits is measured using the trigonometric values between the targeted review representing information and its many past feedback document representations.The similar the reviewing habits of many reviews of the same person and the accompanying emotion scores are, the stronger the similarity. According to experimental findings, document-level sentiment categorization can perform even better when various reviews from the same person have comparable review patterns. For document-level sentiment categorization, LSTM-based models can provide superior document representations, but they do nottake into consideration patterns of user feedbacks.The closer a pair's feedbackroutines are, the elevated the resemblance. We compute the similarities among document representations after acquiring better document illustrationsthat is an important procedure to take into account the resemblancesamong users' review routines. RNN's LSTM variation, nevertheless, can solve this issue. Numerous LSTM-based models can enhance LSTM performance even more.

Shujun Wei and Song Song[40] A multifeatured fusion of text and images-based sentiment analysis technique for online travel evaluations is suggested. In order to finish the text sentiment classification, a text sentiment classification method is first built, and a range of sentiment characteristics are integrated to produce a multi-input matrix. This matrix is then entered into a channel to extract sentiment features. The global picture and the facial image are combined to create an image sentiment classifier model. The supervision modules with weighted loss is added to the CNN base to extract the face emotion features, and the facial target emotion is fused with the sentiment immediately identified by the entire picture, as well as the sentiment polarities of the image in the published tourism review. In order to combine the results of the text and picture sentiment classification models, a decision fusion approach is developed. CNN is used to train and extract sentiment meanings from multi-input embedding's that are more complete. By combining several features, CNN not only creates new features but also enables the features to interact and affect one another.The proposed image-text fusion emotion classification model achieves outstanding results in numerous performance metrics, with better emotion classification execution than some other state-of-the-art models. It effectively improves the model's capacity to capture sentimental semantics of travel reviews through the combined effect of text content features as well as image sentiment features. However, in this work, more complicated circumstances are not taken into account when extracting face emotional features.

M.P. Geetha and D. KarthikaRenuka[41] The use of a tweaked Bert Base unpackaged model was suggested as a way to improve performance of intrinsic part sentiment analysis. BERT is distinctive and different from other machine learning in that it was pre-trained using a sizable unlabeled text corpus that includes a book corpus and Wikipedia. BERT is profoundly bidirectional, from both left to right or right to left text representation. Currently used context-free models typically provide a single embedding for all vocabulary terms, regardless of the context.It has been suggested that the BERT Model may be implemented in two stages: the first stage involves pre-training, during which the model learns to identify the input text data and its context, and the second stage involves fine-tuning, during which the model takes in and recognises the answer. To achieve up-to-date results, the Pretrained BERT model may be fine-tuned by adding one layer. dividing the input review textual information into a predetermined set of tokens in accordance with the glossary. To handle terms that are absent from the glossary, WordPiece tokenization is used. BERT-Base- The model's uncased parameters are assessed. When the validating loss value is low, the best model is constructed, and this is accomplished by adjusting the hyper parameters.The effectiveness of the model is noticeably impacted by changing the hyper parameters. In the empirical assessment, the BERT methodexceeded the alternate machine learning methodsalongside the full understanding and elevated accuracy. To clarify the problem of sentiment analysis, however, this suggested work, the BERT Base Uncased method, a potent Deep Learning method, is provided.

IshaaniPriyadarshini and Chase Cotton[42] created a brand-new grid search-based deep neural network for sentiment analysis called the Long Short-Term Memory-Convolutional Neural Network (LSTM-CNN). The CNN-LSTM model is less effective than the LSTM-CNN approach. The inaugural LSTM layer of the LSTM-CNN architecture is in charge of accepting embedding for each character in the phrases as inputs. The underlying notion is that additional data will be stored in the output token for both the initial and prior tokens. This model's LSTM layer is in charge of creating a new encoding again for original input. The CNN, capable of retrieving local features, receives the output from the LSTM layer.After this convolution layer's output is pooled to the smaller dimension, it is labelled as either positively or negatively. The grid search's primary goal is to identify the best hyper parameters for classifying sentiment polarity more precisely. A fully - connected network, dropout, LSTM layer, pooling layer, grid searching, and output layer are just a few of the layers in the proposed design. The input goes via a multi-unit LSTM layer. This technique's key benefits are that it is simple to use and reliable in terms of search space. Since K-NN makes no assumptions on the data, it is effective at handling nonlinear data. Sentiment analysis is a challenging topic, but advanced techniques may help close the gap between people and robots.

J. Shobana and M. Murali[43] The suggested model makes use of long short-term memory to comprehend intricate patterns in textual material. Weight arguments are modified via the flexible particle Swarm Optimization (PSO)to enhance the LSTM's performance. By modifying features like weights and learning rate to cut losses, the APSO optimizer enacts a crucial part in improving the suggested LSTM neural network model's accuracy. The suggested methodology's objective is to forecast reviewers' opinions using the APSO-LSTM algorithm. Amazon review data is utilized to estimate the execution of the suggested method.In contrast to classical LSTM, APSO-LSTM is better at choosing the ideal weights for neural networks and making wise hyper-parameter selections, which results in increased accuracy. The PSO method is used with OBL to address the optimization problem. For word embedding, the skip gramme method-related extracted features has been employed. When compared to alternative vector representations, the Skip-gram Word to Vector format uses lower memory and consistently produces superior accuracy. The assistance of APSO in LSTM neural network weight parameter selection improves accuracy and reduces computational complexity. Numerous comparison evaluations were accomplished to show the usefullness of the suggested method.Better sentimental prediction tasks are facilitated by wise feature extraction technique selection. However, because the weight arguments of the LSTM are optimisedutilizing the APSO approach, the calculation time of the suggested APSO-LSTM is lowered to that of the LSTM.

Yihao Zhang and Zhi Liu[44] an innovative hybrid recommendation method that combines neural collaborative filtering with paragraph embeddings has been discovered. In order to employ user-item ratings for information retrieval, it uses neural networks. These networks have a high degree of non-linearity, which allows them to seize the intricate form of user ratings. The cold start issue is somewhat resolved while by utilisingproductembeddings to record the context characteristic for secondary data. To represent customer feedbacks and productexplanation, we specifically propose contextual embeddings. We also develop two neural networks to seize, respectively, the sentiment of customer feedbacks and the contents characteristic of products.The proposed hybrid module is designed to provide better-quality recommendations. It unifies a proposed hybrid model by combining one-hot encoding of people as well as objects, text word embedding of user evaluations, and product explanation. Utilize the addition fusion layer in particular to forecast user ratings and circumvent the restrictions of matrix factorization. In order to take use of their complementing features, the hybrid recommendation technique combines two or more existing approaches. However, because it cannot take use of an item's content characteristic, this solution is unable to points the cold beganissue in recommender system.

AytugOnan[45] built a powerful deep learning-based sentiment analysis framework. The suggested framework associated CNN-LSTM framework+ and TF-IDF weighted Glove word embedding. The final design uses this method in the weighted embedding's layer since it produces the best predictive performance: TF-IDF weighted GloVeembeddings with center-related consolidation. Word embedding techniques are used in the hidden layers of the CNN architecture to represent text documents. Pretrained word embedding strategy has been used in this layer because pretrained word vectors provide better prediction performance. For word embedding, a vector length of 300 is chosen based upon the findings. To avoid over fitting, the design adds a dropout layer after the weighted word embedding layer. Convolution layer has been designed after dropout layer. Utilizing the stack of convolution-based properties is done in this layer. To build feature maps, a predetermined set of 80 filter are applied to each layer. Every input text was initially transformed into a concatenated of every word vector. The two other things done are outperformed by those vectors, which are TF-IDF scaled word embedding schemes. The center-based aggregation function performs better than the weighted sum and the Delta rule in terms of the effectiveness of vector aggregation functions.The weighted sum produced the lowest prognostic accuracy, whereas the Delta rule function produced the second-highest predictive performances. relating to how well deep neural network topologies anticipate outcomes. There is no one method that can produce the maximum prediction execution on everytypes of text categorization works, however, and document length can be a limitation when altering the settings of word embedding.

2.2 Product Opinion Analysis using Machine Learning

Due to the rapid evolution of Internet technology, digital shopping has currently turns a common path for consumers to both buy and utilize goods. The computer diligence of spontaneously determining out the sentiments a writer is conveying in text is termed as sentiment analysis. In recent years, sentiment analysis has drawn a lot of interest. Positive and negative are frequently used as a frame, but it may also be more specific, such as defining the precise feeling the author is conveying, such as fear, pleasure, or rage. Business companies may use subject recognition, opinion polarity, and sentiment to determine the causes and the total scope globally. These insights may then be used to expand competitive intelligence, enhance customer service, boost brand perception, and provide businesses a competitive edge. employing web scraping to extract the information from an e-commerce website. The amount of pages or so of remarks for each product will be looped through. Online product reviews were gathered for this project using web scraping. Utilizing classification algorithms and opinion or sentiment analysis, the online product reviews that have been gathered are examined. The categorization model experiments have shown encouraging results.

Huiliang Zhao and Zhenghong Liu[46] designed a new improved Machine Learning (ML) method for the SA of digital product feedback dubbed the Local Search Improvised Bat Algorithm related Elman Neural Network (LSIBA-ENN). e LSIBA-ENN, which categorises the tone of customer evaluations as favourable, unfavourable, and neutral. Two yardstick datasets are used for the performance study of the proposed and existing classifiers. The results show that, when compared to other top algorithms, the LSIBA-ENN succeeds the better execution in SC. The reviewer's views are accurate. It is suggested to use LSIBA-ENN to do SA of digital feedback of products.The data is initially acquired from publicly accessible E-commerce platforms including Taobao, JD, Amazon, and E-bay. The WST is then used to capture or extract text-based consumer opinion data from the websites. Software specifically designed for autonomously pulling data from webpages is called web scraping. Everyone who wishes to obtain detail from a Database in any alternate path can benefit from these resources. It oversees proxy servers and applications that have been severed. With the ability to run JavaScript primarily on users while rotating proxies in response to demand, anyone may access the real Html file without even being hindered. It was frequently used for eCommerce exfiltration since it was easy to use for both coders and non-coders. As a result of the data extraction, the customer reviews of the items are then filtered away. After that, the customer reviews are preprocessed using white tagging, Gensim Lemmatization (GL), and Snow-Ball Stemming (SBS). Tokenization is a method for partitioning a text into bite sized pieces. These bits are sometimes known as tokens.It could reduce a big body of content to a few key words or phrases. Based on the problem at hand, we may establish their precise criterion for dividing the text content into pertinent tokens. But this method could only categorise sentiment into positively and negatively categories, making it unsuitable for applications where sentiment refining is highly required.

Atif Khan and Muhammad Adnan Gul[47] devised a method that associates supervised and unsupervised to classify the feedback as favourable or bad, then to summarise the feedback in the movie review field. Certain that the training set for a given domain are accessible, the recommended technique can be used to any specific domain. The bag-of-words feature retrieval approach is employed to retrieveembeddings, bigrams, and morphemes as a feature set from the feedback documents and illustrates the identifying and analysing potential as a vector for the purpose of categorising movie reviews. The movie reviews are then divided into unfavourable and positive reviews using the Nave Bayes method.For the job of summarising movie reviews, a word2vec method is utilised to retrieve features from categorised movie feedbackstatements. Semantic clustering approach is then utilized to cluster feedback phrases that are semantically linked. To estimate the important score of individual feedback statement in a cluster, many text attributes are used. Finally, a summary of film reviews is created by selecting the top-scoring review phrases based on salience scores. a categorization and summarising technique for movie reviews was suggested. Bag-of-words feature retrieval approach is utilized to retrieve unigrams, lexicons, and morphemes as a feature set from the study and data and present the identifying and analysing potential as a vector for the categorization of movie reviews.The filmfeedbacks (illustrated as a feature vector) are then divided into unfavourableas well as positive reviews using the Nave Bayes method. For the job of summarising movie reviews, a word2vec method is utilised to retrieve characteristics from categorisedfilm feedback statements. Semantic clustering approach is then utilized to group feedback phrases that are semantically linked. To estimate the brightness score of individual feedback statement in a cluster, many text attributes are used. The suggested method may be uncapable to identify useful feature-decision pairings, nevertheless, because grammatical relations cannot be used to determine the semantic link between feature and opinion terms.

SwagatoChatterjee and DiveshGoyal[48] Using machine learning, a healthcare service was created for the e-commerce of healthcare/health-products. In order to reflect and anticipate customer happiness, machine learning and econometric approaches are used to determine which core and enhanced service characteristics as well as which emotions are more significant in various service situations. Econometric method executes correspondingly to the top preferred machine learning methods, according to research on machine learning. Moreover, CSAT may be projectedutilizing the data collected from the qualitative evaluation. As a result, the study provides recommendations for where a health insurance ecommerce organisation should concentrate, along with an automated approach that can quickly identify CSAT reflectors in various service contexts.It is essential for e-commerce enterprises to analyse large data sets, use automated prediction models to suggest possible service designs, and handle client feedback using automated review management solutions. This text analysis and projective machine learning scientific method can instantly retrieve relevant detail from the text and assess the dependent significance of that detail in forecasting consumer delight, furnishing managers with vital marketing data in a dynamic setting, constantly-modifying surroundings.We may draw the conclusion that both the C&A service qualities do have a significant impact on the different types of ecommerce enterprises, especially in terms of reflecting and forecasting CSAT, results of the regression results and the findings from feature significance scores. The MPAA model, where consumers employ several paths when developing attitudes, might be used to explain the aforementioned. However, according to the order-logistic regression, there is no correlation between it with any category's customer satisfaction. For the subcategories of exercise, pharmaceuticals, and skincare, equipment and facilities are crucial.

Shanshan Yi and Xiaofang Liu[49] The Hybrid Recommendation System (HRS) was created using a regression model based on machine learning. This technique has been discovered to be useful in categorising the preferred choice of stores depending on the items that the client has purchased. The most notable aspect of this HRS technique that there is no human factor involved in forecasting client store preferences. At terms of precise customer sentiment prediction with regard to purchasing a product in a specific store, HRS clearly exceeds other modern methodologies. This strategy may be expanded in the future to gauge consumer interest in a variety of items across various regions.A great degree of accuracy is clearly demonstrated by the fact that HRS was almost. The MSE value for HRS also shown minimum variance, which is another potent sign of good precision and accuracy. HRS and its modern techniques. It is obvious that the suggested HRS surpasses other current techniques since the MAE values are noticeably lower when compared to other techniques. But because human emotion is multidimensional, a straightforward binary SVM classifier might not be appropriate for sentiment analysis.

AytugOnan[50] An innovative methodology for classifying and forecasting client attitudes was put forth. Automated classification tree induction using a machine learning technique and training data There are typically two steps in a decision tree training algorithm. Tree growth is the initial stage, during which a tree is constructed by greedily splitting each node. The overfitted nodes of the tree are deleted in the second phase because the tree may overfit the training data. The development of technology and algorithms that enables the deployment of networks with several layers, or "deep learning," is what makes a deep learning tool possible.What matters most is that there be a data collection that can be utilised to feed the machine learning algorithms, regardless of the approach employed. Six variant machine learning methods are utilised in this research to categorise text; these methods are SVM, ANN, NB, DT, C4.5, and kNN. These methods were selected for their usefulness and accuracy in text classification. Using a categorizer provides us to differentiate the outcomes of the six text classification appraoch. These techniques' performance has been assessed by researching their accuracy.Additionally, feature sentiment analysis has been used to assess the consumer voice. Big data may undoubtedly be used to make better decisions, but because they are connected to some of the major difficulties and problems in the area of social media analysis, they do not necessarily result in better marketing.

B. SenthilArasuand B.JonathBackiaSeelan[51] The Waikato Environment for Knowledge Analysis is used while creating a social media marketing plan (WEKA). WEKA is a data mining tool that analyses data and generates the output required for effective marketing. As a result, companies may increase their income and differentiating factor. This type of data analysis provides insight into the purchase habits of customers. Data mining methods may be divided into two primary groups: descriptive and predictive, as well as subdomains. WEKA employs many algorithms for various circumstances, and it has enough algorithms to forecast various market conditions. One of the branches of artificial intelligence is machine learning (ML).In order to recover from and handle business crises, various AI principles are used to meet various market difficulties. The data set to be examined is initially gathered from the internet. Both organised and unstructured social media data may be obtained. The suggested technique preprocesses and transforms the social media data to the appropriate format before doing the analysis. Following the extraction of feature vectors, the collected data is subjected to machine learning operations including classification, prediction, and clustering. WEKA executes superior than alternate tools with regard to employing variant types of mining methods, business applications and data analysis approaches.However, WEKA needswellknownwisdom of database managing.

Anjali Dadhich and Blessy Thankachan[52] presented a Product Comment Summarizer and Analyzer (PCSA) system is an automated and versatile comment analysis tool that effectively determines sentiment polarity. The PCSA system is designed to be generic, robust, and fast, employing five supervised learning classification techniques: Naïve Bayes, logistic regression, SentiWordNet, random forest, and K-Nearest Neighbour. It analyses and categorizes online English comments collected from popular shopping websites such as Amazon and Flipkart. The system identifies sentiments as positive, negative, or neutral by utilizing the aforementioned classification techniques. After conducting a thorough review and survey, it was observed that the PCSA system leaned towards positive comments. The need for the PCSA system with multiple classifiers was reinforced by the comprehensive research. The PCSA model consists of two stages: the training stage and the testing stage. It utilizes two storage media, namely the Natural Language Tool Kit-based Corpus and class repository. The NLTK-based corpus comprises an English dictionary, a list of stop words, and WordNet information. The PCSA system is an automatic and versatile comment analyser capable of effectively identifying sentiment polarity. However, despite its extensive collection of product reviews, there is still a demand for an efficient sentiment analysis classification system that can handle multiple online products from various data domains and sources without compromising accuracy.

Lei Li and ShumingRong[53] offered a summary of drinking water treatment (DWT) approaches using artificial intelligence (AI). The administration and administration of DWT processes are supported technically by AI technology, which is more effective than depending simply on human operations. AI-based analysis of data and adaptive learning mechanisms have the capacity to provide a platform for process modeling and predictive modelling that can diagnose water quality, make autonomous decisions, and optimise operational processes. This article provides a quick introduction to AI technologies utilised often in DWT.Furthermore, this paper reviews in detail the established uses and most recent developments of artificial intelligence (AI) and machine learning (ML) technologies in the areas of source water quality, clotting cascade, disinfection, and membrane filters, such as source water contaminant surveillance and identification, precise and effective prediction of clotting dosage, analysis of the creation of disinfection by-products, and sophisticated control of membrane fouling. Since the AI technique is not dependent on any assumption under ideal circumstances, it is more suited for simulating the treatment of water under actual conditions.Its setup procedure is more streamlined and quick, and it places a stronger emphasis on predictability than merely model fitting. More powerful thanks to the ongoing development of AI technologies and the use of integrated AI. The benefits of one strategy are combined by AI. More crucially, even with only two input factors, the big data model still produced predictions with a reasonable level of accuracy. However, there are still certain difficulties with applying AI technology in DWT. The current interpretable analysis techniques concentrate on local analysis; further investigation is required to determine how the water purification problem and the neural network structure are coupled, as well as how to acquire the dynamic of neural network enhancement under the coupling principle.

Mohamed Elhag Mohamed Abo and,NorismaIdris[54] created a multiple-principles strategy to evaluate and rank Arabic sentiment analysis classifiers empirically. This study's main goal is to present a multi-criteria strategy for choosing a suitable machine learning approach for sentiment classification of dialect Arabic and to illustrate it on a corpus of Saudi tweets. To choose the best classifier for vernacular Arabic, we distinguished the execution of five machine learning techniques categorizers: learning based, decision tree, Naive Bayes, K-nearest neighbours, and support vector with multi-criteria. The experiment's findings demonstrate that deep learning and the support vector machine classifiers outperformed decision trees, K-nearest neighbours, and Naive Bayes classifiers with regards to effectiveness, precision, recall, F-measure, and AUC. The suggested technique was used to sentiment labels obtained from web tools. Additionally, many students were encouraged to sign up for the online system and categorise the texts composed of brief review phrases as good or negative, which included 5255 positive or 6392 negative tweets.It may be used to help both newcomers and established academics better understand the theories and ideas behind machine learning techniques, performance metrics, and sentiment classification of dialect Arabic texts in order to find a long-term solution to its morphological centered problems. Convolutional and support vector classifiers are not necessarily the best techniques, though, since several other algorithms with different datasets showed promise.

MehrbakhshNilashi and HosseinAhmadi[55] created a novel strategy using machine learning for consumer segmentation and preference prediction. Text mining and predictive learning techniques were used to construct the system. Customer segmentation also used a clustering approach. The approach was assessed using information gathered from vegetarian-friendly Bangkok eateries. developed a machine learning method based on TripAdvisor data for hotel categorization and vacation selection prediction. Customer feedback gathered using a crawler that uses PHP scripting. Collect information first from TripAdvisor, a social networking website. Data was gathered using a specialised crawler.Following that, we preprocess the data. This process is used to convert raw data into a format that can be understood. Additionally, in this stage eliminate the quick reviews that did not contain pertinent data for knowledge extraction. Additionally, reviews written in languages other than English and those without a numerical rating were eliminated. The approach is created utilising CART for preference prediction, SOM for data clustering, and LDA for text mining. The prediction methods in CART may be learned as regression trees using quick, dependable techniques. In addition, CART is capable of managing big datasets with incomplete data. But using binary decision trees constructed from the predictor variable, this method solves the classification or regression issues.

Soumya S and Pramod K.V[56] This work developed a Sentiment Analysis of Malayalam Tweets using Machine Learning Approaches. Using several machine learning approaches involves Naive Bayes (NB), Support Vector Machines (SVM), and Random Forest, the tweets are divided into positive and negative classifiers (RF). For the purpose of creating the feature vector for the input dataset, several characteristics like the Bag of Words (BOW), Word Frequency vs. Document Term Frequency (TF IDF), Unigram with Sentiment lexicon, and Unigram with Sentiwordnet adding negation words are taken into consideration. Based on terms in Malayalam with a focus on both positive and negative emotion, the Twitter API was used to get 3,184 tweets.Positive tweets that have been recovered occasionally display bad sentiment, and vice versa. Therefore, each and every one of the received tweets is carefully reviewed and given its true emotion. Due to the importance of sentiment-oriented words in predicting the sentiment of sentences, all three classifiers using Unigram with Sentiment lexicon and Unigram with Sentiwordnet incorporating negators perform superior. However, as they did not exclude irrelevant terms, the feature vector was extremely large.

Arshad Ahmad and Chong Feng[57] The main objective of this Systematic Literature Review (SLR) study is to recognise or identify and categorise or classify the types of machine-learning algorithms or methods used for identifying software components on the Stack Overflow platform. This goal was implemented thoroughly during the planning, conducting, and affects the productivity of the SLR work. Our SLR research is to uncover several machine deep learning or strategies that have been effectively applied to determine the various software needs on the So, as well as their functioning and assessment processes.The results will finally assist and enable us to identify the primary challenging concerns and challenges that must be correctly addressed for the purpose ofenhance the functionality of the various machine learning-based approaches. While taking into account the significance, organisation, and amount of the text, as well as the performance of both these ML algorithms, the findings were adequate. SLR did not, however, precisely state why or how crucial it was to use this metric in their research. Only that this method is frequently utilised for information extraction tasks was asserted.

Mohammed Ali Al-Garadi and Mohammad Rashid Hussain[58] identified a substantial body of research to use machine learning techniques to detect hostile conduct on Social Media(SM) platforms. An whole new type of anger and violence that takes place online has been introduced by SM platforms. This research highlights a novel method for displaying aggressive behaviour on SM websites. The reasons for developing prediction models to counteract aggressive behaviour in SM are also discussed. Millions of users worldwide utilise SM websites as dynamic social communication platforms. Through online social contact, information in the form of ideas, ideas, preferences, viewpoints, and conversations spreads quickly among users.SM users' online activities produce a vast amount of data that may be used to analyse human behavioural patterns. In order to forecast aggressive behaviour, SM websites are used. Such an analysis for predicting cyberbullyingbehaviour is restricted to textual OSN material. Given how simple it is to engage in cyberbullying, it is regarded as a hazardous and quickly proliferating hostile behaviour. Bullies just need to be ready to act inappropriately and have access to the Internet on a laptop or mobile device to commit misbehaviour without approaching victims. These research presupposed that SM sentiment traits serve as a reliable indicator of the incidence of cyberbullying. Traditional approaches are difficult to scale and accurately use in SM in this situation.

M.R. Martinez-Torres and, S.L. Tora[59] used a review-centric methodology and identified the polarity-oriented distinctive properties and polarity-oriented distinctive subjects for false and honest feedback. Four sets of words with effective discriminant qualities between classes are provided by the suggested polarity-oriented of unique attributes. They may be represented visually using correspondence analysis, a grouping technique that has gained popularity for dimension reduction and sensory mapping. It is used to discover similarities and associations between variables. In order to identify the distinctive characteristics and subjects of deceitful and honest feedback, this work advances the area of review-centric methods by utilising a number of machine learning algorithms.The main conclusions of this research are the detection of linear polarization own characteristics that can distinguish between misleading and non-deceptive reviews based on their polarity alignment and the identifiers of topics for deceiving and honest feedbacks while also taking their antagonism introduction into account. Extreme opinions are frequently expressed online as a result of very positive or negative experiences. In these situations, honest evaluations might be just as favourable or bad as those that are assumed to be dishonest. However, because bi-grams often occur far less frequently than unigrams do, it is unlikely that the inclusion of bi-grams will have a significant influence on the outcomes.

DoniaGamal and Marco Alfonse[60] generated machine learning (ML) methods that are used in the sentiment analysis and review mining across various datasets. guided, unsupervised, and semi-supervised learning, which is an ML method that is often employed in sentiment analysis (SA). In supervised methods, the data is labelledfor the purpose ofattain an accurate and logical outcome. Unsupervised learning is not needslabelled data, unlike supervised learning. Clustering approachareutilized to handle the problem of processing unlabeled data. In order to evaluate the effectiveness of several machine learning (ML) algorithms for text categorization and compare their assessment accuracy, four distinct datasets were used in this study.With regards of successfully emply ML approach to the SA datasets, it is crucial to retrieve reliable textual characteristics that result in successful right categorization. A five-step system has been developed for SA. First, a dataset is used, and so on. A five-step system has been developed for SA. A was chosen in the first phase from one of the four separate datasets. The second stage involves pre-processing, and a dataset is chosen from among the four available datasets. The second phase comprises pre-processing, while the third stage involves applying FE methods to the chosen dataset. The following phase is an all-ML process, while the third stage involves running FE algorithms on the chosen dataset. Algorithms are then trained. Finally, 10-fold evaluations of the various ML methods are performed. All ML techniques have been trained. Finally, 10 folds are used to compare the various ML methods. Bigrams and morphemes surpass unigrams in the different reviews dataset, according to various analysts and academics, whereas ML outperforms bigrams in categorising film reviews by sentiment polarity.

Ivens Portugal and Paulo Alencar[61] developed a comprehensive literature review that highlights research possibilities for software engineering and assesses the application of Machine Learning (ML) methods in recommender systems. Computers are used in machine learning (ML) to replicate human learning, detect and educate from the real world, and improveexecution on specificworksutilizing this new detail. The goal of the machine learning method is to educate from training data and then employ that wisdom to actual data thatmight be a table whichconnects details about individual book to the appropriate classification. Here, details on individual book may include its author, title, or all word also.With the training data, the ML algorithm gains knowledge. The algorithm can categorise a new book using the knowledge it has learned about book categorization as it comes into the bookstore. It is challenging to choose an ML algorithm for usage in RSs. Researchers in RSs also lack a clear understanding of the utilisation patterns for ML algorithms, making it difficult for them to prioritise their research projects. Software engineering (SE) is a discipline that researches all phases of computer software creation, from conception through maintenance. The field provides resources that can help with the creation of RSs with ML algorithms.Researchers must understand which SE region of RS growth needs resources, nevertheless. However, semi-supervised ML techniques can also be categorised. Algorithms use semi-supervised learning when they must still learn from a training set that contains missing data.

Abhilasha Singh Rathor and AmitAgarwal[62] devised a study to estimate the effectiveness of three machine learning approach. For the categorization of digital feedback utilizing a web methodutilising supervised learning, Support Vector Machines (SVM), Naive Bayes (NB), and Maximum Entropy (ME) were used. Sentimental Analysis is resolved utilizing machine learning (ML) methods that rely on reputed ML approaches as piece of a normal text classification issue that makes use of syntactic and linguistic variables. Amazon review sites are crucial for both the firm selling the goods and the individual customer who purchases it. The research presented in this paper demonstrated that review sites are a salient source of detail for enterprise that vend goods and for customers who are evaluating the option to buy. Textual item feedback should beinitiallycategorised as favourable, negative, or neutral because of their unstructured nature. For accurate prediction, the effectiveness of user evaluations needs to be increased. The suggested method theoretically outperforms Naive Bayes. However, Naive Bayes (NB) classifier works well in practise on a variety of issues. However, the suggested research found no significant differences in categorization at the sentence or document levels.

MonalisaGhosh and GoutamSanyal[63] created a feature selection strategy that was both separately and together ineffective on four classification machines algorithms. To overcome the issue of sentiment classification, two approaches were used: lexicon-based and machine learning. Based on training and test data sets, the former method was used to categorise the attitudes. The second category completes the task by finding a list of words or phrases that has a semantic value; it does not require prior training data sets. It mainly focuses on patterns in previously unobserved data. Using machine learning techniques, you may read feelings in text about how society has influenced a student's life. They used the k-means technique to cluster data, determined the key impacts, and then judged those to be at the class level.The main goal of this research project is to examine how different Machine Learning Classifiers (MLC) perform when three feature sets are mixed. Collection of data, pre-processing, feature selection, and categorization are the four steps that make up the entire process. In compared to the results achieved by using each individual feature selection technique, the adopted strategy that combines several feature selection approaches offers a better outcome. But in other reviews or comments, users conveyed their feelings using emoticons or graphics; we haven't taken these forms of emotions into account for analysis.

Jun-Ho Huh[64] Based on a person's hobbies, degree of attention, and body type, the big data analysis method for tailored health activities that is described in this work is developed. Big data analysis is a technique for displaying the study's main terms, and word clouds help people grasp concepts and keywords in documents more easily. For instance, there is a method that makes it possible to explain a word as fully at a look as it is when it is uttered. When analysing big data, which deals with a vast volume of information, it is mostly utilised to determine the features of the data. R Studio, a big data analysis tool, offers several packages for crawl, text analytics, and word. Big data uses keyword analysis for tailored health activities to analyse search data from a Korean portal firm and Google, as well as unorganized citizen health data from Korea. With the use of text mining and word clouds, the huge data was displayed. This study gathered and evaluated information on obesity-related interests, changes in those interests, and treatment-related publications. As a result, this study used a crawling technique to collect big data, a text mining method to visualise big data, a text and word cloud to display big data, a machine learning approach to assess individual healthbehaviours from various angles.This strategy is used to provide metrics for activities that promote customised health from several angles. A frequency table can be used to combine data to more clearly display the general features than raw data alone. However, because big data analysis is so complex and challenging, it can be challenging to understand its purpose and direction at times.

Zhiyuan Chen and Le Dinh Van Khoa[65] It seeks to offer a thorough overview of machine learning technologies used to identify suspicious transactions and anti-money laundering AML algorithms. The following crucial attributes should be included in AML solution instruments used to spot suspicious transactions: data quality, high detection, scalability, and reaction speed. In the AML industry, the domain specialists carefully hand-pick the attributes utilised to train the algorithms for machine learning. The majority of literature frequently lists which characteristics or features are employed for a certain deep learning training, but no thorough explanation of the feature selection procedure is given. Due to the different dimensions of banking transaction data, an automatic keyword selection procedure is crucial.A state-of-the-art AML system must be able to manage the discovery of new unusual occurrences from unseen transactions and the prediction of unseen instances without lowering the precision of the detection performance in order to maintain the detection rate accuracy and lower the number of false positives. However, it remains challenging to give advice based on logic regarding whether banks should provide decision- management system based on the business rules software for AML.

2.3 Problem Statement

- There is absolutely no need for manual intervention in the deep learning-based technique. Through to the neural network structure, it can automatically choose and extract features and learn from its mistakes.
- However, the experimental data could not be used to total the fluctuations in the classifier due to modifying the phrase length.
- A potent Deep Learning method termed as BERT Base Uncased Models is provided to clarify the problem of sentiment analysis.
- Since bi-grams typically occur less frequently than unigrams do, it is unlikely that their addition will have a significant impact on the outcomes.
- Traditional methodologies are difficult to scale and accurately use in SM in this situation.

2.4 Summary

Today, sentiment analysis is an essential method for gaining insight into opinions, attitudes, and feelings concerning various people, things, products, and services. Researchers have been developing a method to analyse data from social media and social chains for the last several years in order to extract hidden information from them and use that knowledge to create meaningful patterns and conclusions. Sarcasm is one type of emotion that, according to sentiment analysis, comprises the reverse of what you truly want to convey. Humans use it to belittle or tease others. Sarcasm may be used to be funny and demonstrate foolishness. Sarcasm can be communicated vocally or by particular gestures, such as raising the eyebrows or rolling the eyes. There are several methods used to identify sarcasm. In this work, we attempt to discuss the most recent and popular methods for sarcasm detection. This study compares the outcomes of several vectorization techniques used to ML and DL methodologies. This will provide the researchers the ability to select the ideal vectorization method depending on the supplied dataset. Various mechanical learning and deep-learning algorithms are used to the static data and live tweets to perform sentiment analysis and spam identification.

CHAPTER – III

STUDY ON PRODUCT OPINION ANALYSIS FOR CUSTOMER SATISFACTION ON E-COMMERCE WEBSITES

E-commerce platforms have evolved with a wide variety of advertising benefits for consumers to post or share their experience with the obtained goods by submitting reviews that include beneficial remarks, thoughts, and opinions on the product. Nowadays, a huge number of customers has the ability to compare products in online stores and select their top selections in computerized merchants, for instance, Amazon.com as well as Taobao.com. Analysis of clients in electronic platforms & electronic commerce site contains critical electronic information about products. Sentiment Analysis (SA) is widely used as the speaker of the client in applications that aimed towards marketing as well as customer service. Sentiment extractors, in their most basic form, categorise communications as having a positive, negative, or occasionally neutral premise. A common use of sentiment analysis is the automated determination, if digital feedback comprises a favourable or pessimistic assessment. As a result, the words relating to certain features are first identified from online item evaluations using sentiment analysis algorithms. Here, the Deep Learning (DL) approach is utilised as a categorization method for displaying the feedback status. The results provided a proposed location for the customer based on early reviews, previous reviews, and the client's response to the inquiry review. Furthermore, it is observed that the provided method is capable of responding to all of the feedback with a better connection as a human response to the customer.

3.1 Introduction

With the dramatic rise and advancement of community, online shopping has evolved into an essential component in meeting people's everyday consuming needs [66]. Client satisfaction is a mental condition resulting from their personal view of an item or service depending on their expectations along with real performance. Client satisfaction is important for businesses that it allows for service or product enhancement, market analysis, especially customers' behavioural attachment [67]. Client satisfaction surveys have generally been used to assess the critical features, or features, of total customer satisfaction [68]. As a result, unstructured big information through online reviews has both harmful and beneficial effects on customers [69]. Because of the fast rise of electronic commerce, research on online evaluations has become a well-established area [70]. It is vital to investigate how item qualities impact the satisfaction of customers in order to increase it [71]. More precisely, the consumer has a positive perspective towards "quality" as well as a negative perspective against "price." which are known as aspects [72]. Analysing such opinion & sentiment data is becoming increasingly crucial including service, product suppliers and consumers, since it influences consumer purchase choices [73].

Sentiment Analysis is an approach in Natural Language Processing (NLP) that assists in the detection of feelings that might enable companies to collect data about their customer's opinions through various online platforms such as social media, polls, e-commerce website evaluations, and so on. This data helps to understand the causes of product degradation and the factors that influence it [74]. SA is used in the market data, which including evaluating user satisfaction regarding products or fixing their deficiencies, projecting price adjustments based on news attitudes, generating innovative goods / processes, marketing and upgrading items based on client feedback [75]. There are various researches that investigate the economic results and the causes of customer reviews while making recommendations for the design of product review systems, such as providing reviewers with a predefined review structure. Despite the fact that, a large amount of review system configuration highlights has been offered throughout time, few have really been examined. Innovative online action plans and circumstances, are the two-sided stage organisations have emerged. These enable two-sided reviews thus necessitate changed plan elements, for example, to reduce connection in two-sided review systems. Furthermore, most of the system features have been studied for fixed devices such as PCs. Regardless, internet reviews are increasingly being provided and deleted using cell phones, necessitating particular schemekey points. These are the unresolved challenges in ranking research for digital feedback, as well as numerous research studies are still looking into them.

3.2 Problem Statement

- Most of the individuals check internet reviews on a regular or irregular basis prior buying the product. As a result, early reviews on before making a purchase have a significant influence on the success of product sales.
- The quality of the internet sites was not dependent on the quality of the products offered by physical shops, and the metrics were inapplicable for evaluating product categories.
- The customer remarks were based on the desired stock availability of the clients that were not applicable throughout the product's design and development.

3.3 Contribution

- In this research, the online reviews are collected from a raw dataset named as Amazon which contains roughly 142 million product reviews collected between May 1996 and July 2014.
- The collected reviews are given as an input to the process of text extraction which improved the review quality of collected product.
- Next, the process named as Text pre-processing is performed which accepts the extracted text as input from the above process and produces cleaned tokens.
- Latent Semantic Analysis (LSA) is employed in NLP in the process of semantic analysis to evaluate the relationship between a group of files and the words.

3.4 Proposed Method

In this research, a deep learning model is proposed for estimating the best products in various e-shopping websites which including Flipkart, Paytm, Amazon, etc. by assessing the product reviews posted by the customers. At an initial step, the various products datasets of specific organizations are collected from the E-commerce websites. Each product has unique characteristics that allow it to be classified based on a certain quality. This includes aspects like considerably influencing customer reviews, readability, subjectivity, length-significantly, as well as emotion polarity. For collecting FCM model, the clear customer surveys that updated items are created. Figure 3.1 depicts the flow diagram of the present research.

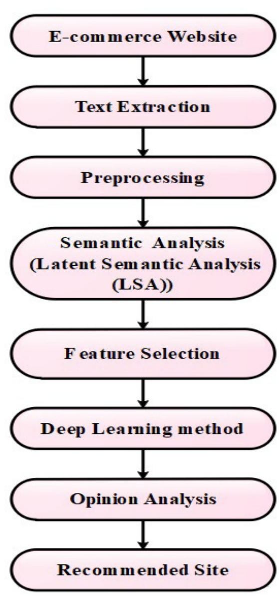

Fig. 3.1 :Overview of proposed Model

3.4.1 Dataset

A raw dataset that is attained from the various E-shopping websites which including Flipkart, Paytm, Amazon, and so on. Even in huge population, these kinds of websites are more crucial in creating an efficient way to purchase their products. The massive trend of such websites is recognised by many administrations, which support in the contribution of different items from multiple categories. Amongst an E-commerce database included for this study is the Amazon dataset, which contains roughly 142 million product reviews collected between May 1996 and July 2014. Every review on the site consists of a written comment submitted by a consumer, which is coupled by an actual time stamp in the study. Typically, reviews are connected with a five-star ratings system that is paired with a written description.

3.4.2 Text Extraction

The utilised text extraction method extracts the customer reviews from the given dataset which the improved the review quality of collected product from the E-shopping websites at the time of analysis. The product improvement for completing audit rating of every product is included in the text extraction. The texts here states to the price, quality, as well as good survey results. The text extraction analyses the opinion of consumer reviews, summarises the reviews, stores the reviews, and prepares the step.

3.4.3 Pre-Processing

It is a difficult procedure to convert words into anything, while, the NLP supports many of the text preparation as well as pre-processing approaches. Text pre-processing accepts extracted text as input from the above process and produces cleaned tokens. Tokens are words or sets of phrases that are counted by frequency then employed as analytical features. The following are the process related with text processing in this section.

- Tokenization,
- Removal of punctuation
- Stop word removal
- Lemmatization

3.4.3.1 Tokenization

Tokenization refers to the process of transforming a constant flow of text into words, phrases, symbols, or even alter relevant components termed as tokens. The aim of tokenization is really to search the terms in a phrase. Tokenization is important in linguistics as well as computer science, as it is used in tokenization. Textual corpus is just a string of characters at first. The keywords in the set of data are acquired for every steps in detail extraction. As a result, the need for a scanner is document tokenization. This may appear to be a minor issue because the text has already been saved in computer forms. Tokenization is mostly used to detect important terms, in this work, the online reviews are tokenizing into words.

3.4.3.2 Removal of Punctuation

Punctuation contributes around 40% to 50% of the words in a hard copy. A certain sentiment analysis model's output is unaffected by punctuation. Even though there is no influence on the sentiment analysis, the punctuations must be removed. All the punctuation from word is deleted at this point then offered the information in its normalised form. The resulting text is condensed & simplified and the file is edited to remove any punctuation.

3.4.3.3 Removing Stop Words

Many terms in texts appear repeatedly but are still basically useless since they are applied to link words in a sentence. Quit words, as it is frequently considered, may not help to the content as well as substance of written writings. Due to their usualinstant, their reality in text mining makes an impediment to understanding the substance of the texts. Quit words are frequently employed basic phrases which including 'and,' 'are,' 'this,' and so on. They are impact less for document categorization. As aout come, they should be removed. However, developing a certain list of stop words is challenging & inconsistent among literary sources. This procedure also minimises text data while improving system efficiency. Each written document handles with these terms that are not in the dictionary. This procedure also minimises text data while improving system efficiency.

3.4.3.4 Lemmatization

Lemmatization is a key pre-processing stage for several text mining applications. It's also utilised in NLP as well as a variety of other domains related to linguistics. It may also be used to produce general keywords for browsers or labels for idea mappings. Lemmatization is identical to word stemming because it requires replacing the suffix of a term occurring in textual information with a (usually) different term suffix to obtain the normalised word form. It is thought to be an additional method for dealing with affectation by selecting on the grammar rules & employing a point-by-point dataset of the languages.

3.4.4 Semantic Analysis

The semantic analysis is conducted on pre-processed data, in which the natural language content scans each and every words then catches the information by executing the actual significance of every text. The text components are determined depends on their logical as well as grammatical purpose. The semantic analysis examines the adjacent contexts in the document to determine the precise interpretation of each word. The connection among the ideas in the text is also formed in order to determine the most crucial aspects in the text as well as analyse the issue presented. Latent Semantic Analysis (LSA) is employed in NLP in the semantic analysis to evaluate the relationship between a group of files and the words that are include by constructing a collection of ideas connected to the files as well as words.

3.4.5 Feature Selection

A feature selection method is done on the extracted concepts to choose the subset words. Those subset words are obtained in training, & all those chosen subsets are considered as features which conducted the process named as text classification. Initially, training is done and classifier is used successfully to reduce the amount of the vocabulary. Furthermore, feature selection increases accuracy and hence reduces noisy characteristics.

3.4.6 Deep Learning Approach

Deep Learning approaches offer a solution to obstacles in NLP issues which including sequence-to-sequence prediction. The constructed model utilised DL approaches to learn the features needed by the model then those features are extracted. DL performance in NLP is dependent on actual outcomes and that the upgrades appear to be progressing & possibly speeding up.

3.4.7 Opinion Analysis

The ideal sorting viewpoint evaluates the importance of every aspect in relation to the sentiment score used to calculate sort key points. The commenters assessed the sites that have crucial facts as well as differentiated the untruthful viewpoints related to opinion analysis.

3.4.8 Recommended Site

Depending on the opinion analysis, a viewpoint positioning computation is done for ranking the key points, which resulted in a perspective repetition which influenced views of every perspective for general feelings.

3.5 Results and Discussion

A new opinion analysis framework is created for precise recommendation of products by assessing the product reviews posted by the customers. The major motive of the current work is to develop a most precise keyword extraction approach as well as the clustering technique for suggesting the product in both positive as well as negative kinds by utilising the dataset named as amazon customer reviews. In this research, by employing GWO algorithm, a keyword extraction approach named as Latent Dirichlet Allocation (LDA) which assist in the selection of accurate keywords. The PCFM algorithm is used for clustering the similar keywords that are obtained repeatedly. Moreover, the created recommendation system has many useful benefits which including, the created system has the capability of detecting the fake products and it keep monitoring the customer's satisfaction. Table 3.1 shows the comparison analysis between the proposed as well as traditional approaches in terms of various performance measures. Figure 3.2 depicts the graphical representation of comparative analysis.

By performing quantitative as well as comparative analysis, the proposed model produces a well-defined result. From the test analysis, the proposed model achieved an Accuracy of 77.236, AUC of 57.404, F1-measure of 74.167, Recall of 76.781, as well as Precision of 76.106. Whereas, the traditional approaches such as Decision Tree (DT), Random Forest (RF) obtained limited results on amazon customer review dataset which are given as follows: Accuracy of 75.528 & 73.187, AUC of 54.628 & 48.263, F1-measure of 74.618 & 71.132, Recall of 75.583 & 73.25, Precision of 75.766 & 69.003 respectively.

Table 3.1 :Comparison between Proposed and Traditional Approaches

	DT	RF	Proposed
Accuracy	75.528	73.187	77.236
AUC	54.628	48.263	57.404
F1-measure	74.618	71.132	74.167
Recall	75.583	73.25	76.781
Precision	75.766	69.003	76.106

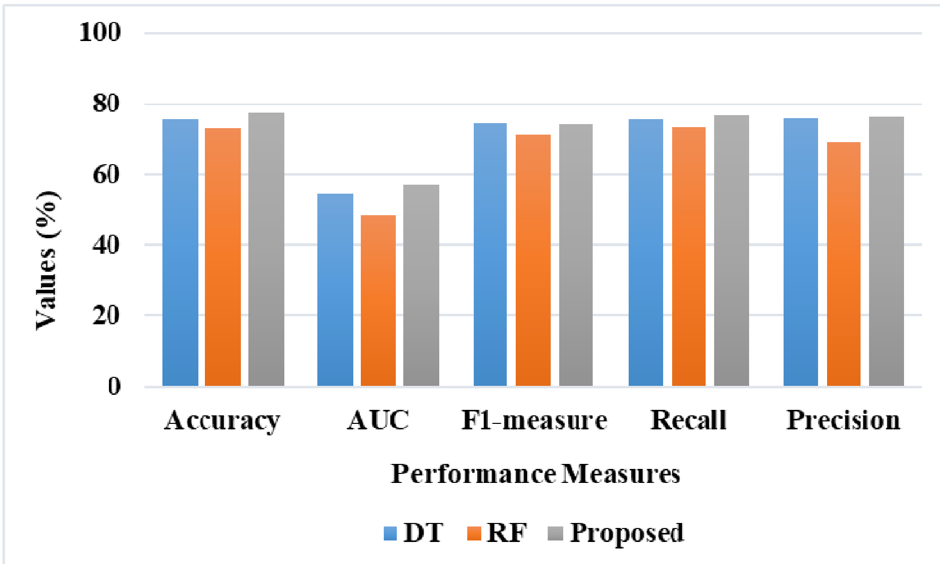

Fig. 3.2 :Graphical Representation of Comparative Analysis

3.6 Conclusion

In recent times, the development community highly concentrating on the user's browsing experience, due to the amount of individuals utilising the internet is increasing at an exponential rate. This research offered a method for predicting the most effectual online shopping sites. A client survey has evaluated that if the product is excellent or unsatisfactory from business websites. It is significant for any kind of firm to be aware of customer feedback on any items. Amazon datasets is used in this research to evaluate the features then categorise early reviewers on e-commerce sites, as well as their influence on product popularity. Various process which are included in this research such as text extraction that improved the review quality of collected product. The outcome from this process is given as an input data for the pre-processing step which produces cleaned tokens. Then, the LSA is employed in NLP in the semantic analysis to evaluate the relationship between a group of files and the words. A feature selection method is done on the extracted concepts to choose the subset words. This work proposes to analyse the posting procedure and construct a DL model for predicting reviewers. Depending on decision research, a suggested viewpoint branding calculation to sort the basic statements by considering the repetition of viewpoints as well as influence of views which yields a recommended site.

CHAPTER – IV

WORD EMBEDDING ATTENTION & BALANCED CROSS ENTROPY ECHNIQUE FOR SENTIMENT ANALYSIS

Nowadays, a lot of data like raw facts, reviews, or opinions has stored on social media or e-services websites. In recent years, people have preferred to purchase products in online, so, the customers choose the best products and a lot of information is gathered in the form of customer feedback. Customer feedback helps with product development and boosts sales of the product.Reviews posted by users on e-commerce websites offer useful details about the product.Sentiment analysis on the text review helps to analysing the user sentiment toward product and predict sales of the product.The Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN)-based methods utilized in current methodologies for sentiment analysis involve the vanishing gradient problem and overfitting problem, respectively. For the purpose ofenhance classification execution, the Word Embedding Attention (WEA) strategy is introduced in the Bi-directional Long Short Term Memory (Bi-LSTM) model. From the input dataset for sentiment analysis, the CNN model is used to retrieve the features. To preserve the gradient in the network and resolve the vanishing gradient problem in the network, the balanced cross-entropy is presented.The WEA technique gives terms with a strong connection to class higher weight, this method helps to improve the model's execution with regards to classes, which increases an accuracy and recall value.The CNN function helps to improved performance with less training data.While the existing CNN model has 97.1% accuracy and 85.4% precision in sentiment analysis, the WEA-BiLSTM model has 97.4% accuracy and 86.8% precision respectively.

4.1 Introduction

The COVID-19 pandemic and recent developments in internet technologies have advanced digital transformation and increased usage of social media and e-commerce platforms. Online consumer reviews are a common and expanding form of User-Generated Content (UGC) and are maintained by a number of review websites, including Amazon, Yelp, and TripAdvisor. Before making a purchase, customer’s check the feedback to evaluate the high standard of an organization, item, or provision. Online customer reviews generally have a significant influence on many prospective consumers' shopping decisions [76]. Due to the fact that consumers prefer to buy things online, a lot of information is gathered in the form of customer feedback to help potential buyers select the best items. These studies provide optional languages that helps the e-commerce industry identify areas that require development.

The information is very useful for businesses looking to comprehend customer feedback on their goods or services. The emotional analysis data is really useful for companies seeking to comprehend customer feedback on their goods or services, as well as experience [77]. Sentiment analysis is the technique of identifying the intended sentiments that are expressed in text-based resources like social media postings, product reviews, and online communities. The majoraim of sentiment analysis is to make it easy for machine learning-based algorithms to recognise these intended emotions when they are conveyed in textual materials. The application of sentiment analysis has performed highly effective in a variety of fields. Similar to this, companies typically use the sentiments collected from client reviews and comments to polish and enhance the USPs of their goods [78].

One of the most essential concepts in Natural Language Processing (NLP) and text mining for sentiment analysis is text categorization. Associating pertinent material with labels that already exist is the description of the text classification problems. In this rank of labelling, set of data form is important. Each text can be illustrated by a single label or several labels, based on the issue's current state. Stock trading firms also collect information and assess their view of various facts using sentiment analysis technologies [79]. Using sentiment analysis, characteristics of software products may be extracted from user evaluations together with a summary of what users think of each feature. Product managers could find this information quite helpful in prioritising their work for upcoming releases. It is also possible to utilise the user assessment at the feature level to sporadically check on the features and general health of the product. Review classification therefore emerges as the fundamental tool for textual data organisation [80]. Based on user feedback, product characteristics may be categorised into many groups. The quick advancement of machine learning and artificial intelligence opens the door to the computer analysis of customer evaluations for opinion mining. Customer review data offers various benefits, starting with the fact that there are lot of reviews that are publicly available online. Second, there are several extremely useful qualities that may be used to characterise customer review data. Third, customer reviews are raw, open data that are collected from regular people and made available to all manufacturers [81]. In essence, consumer feed back display how the customer feels about the enterprise that is important to understanding what the consumer thoughts. Theseestimation significantly influence other customers' choices and serve as the cornerstone of business development [82] without any client feedback on their goods or services. Customers can offer feedback by way of an online survey, a social media website, or handwritten evaluations. There are a tonne of thoughts and evaluations available, but it might be challenging to classify them as favourable or unfavourable. Customer satisfaction is crucial to the hotel industry's ability to deliver higher-quality services and improve customer relationships [83]. However, the fast growth of online purchasing has resulting in an excess of product reviews, making it challenging for customers to choose the most useful ones. Therefore, it becomes crucial for platforms to effectively identify useful reviews in order to help consumers make intelligent purchasing decisions. As a result, large reviews sometimes include a lot of pointless or boring material, which lowers the categorization performance of review length. [84]. As a result, the purpose of this study is to determine whether customer feedback from applications and online reviews is related. Customer sentiment is the term used to describe the sentiments that consumers express through text reviews. These feelings can be either good, negative, or neutral. The research investigates consumer feelings and describes them in terms of the polarity of those ratings [85].

4.2 Problem Statement

- The LSTM model has vanishing gradient problem and HNN model has overfitting problem in classification.
- The fusion of features creates overfitting technique and irrelevant features affects the classification performance.
- The learning capacity of the model was less and the feature selection process was not effective.

4.3 Contributions

- The WEA technique is proposed to provide higher weight value to the words having strong relation to the classes. This helps to increases the learning rate of the method and expands the class wise execution of the method.
- The balanced cross entropy is applied to maintain the gradient in the network to reduce the vanishing gradient problem. The gradient is maintained in the network based on the loss calculation in the network.

- The CNN based feature extraction technique is applied to generate more features in the convolutional layer and helps the model to provide higher performance for less number of training data.
- The WEA-BiLSTM model has higher performance in sentiment analysis compared to existing techniques. The WEA-BiLSTM model solves vanishing gradient problem and reduces the overfitting problem in the classification.

4.4 Proposed Method

The CNN model is used to retrieve features from the input dataset of Amazon reviews. The Bi-LSTM model's attention layer applies weight values to the features in order to classify sentiment. The network's gradient is maintained by applying the balanced cross entropy. The WEA-BiLSTM model's process for categorising sentiment is shown in Figure 4.1.

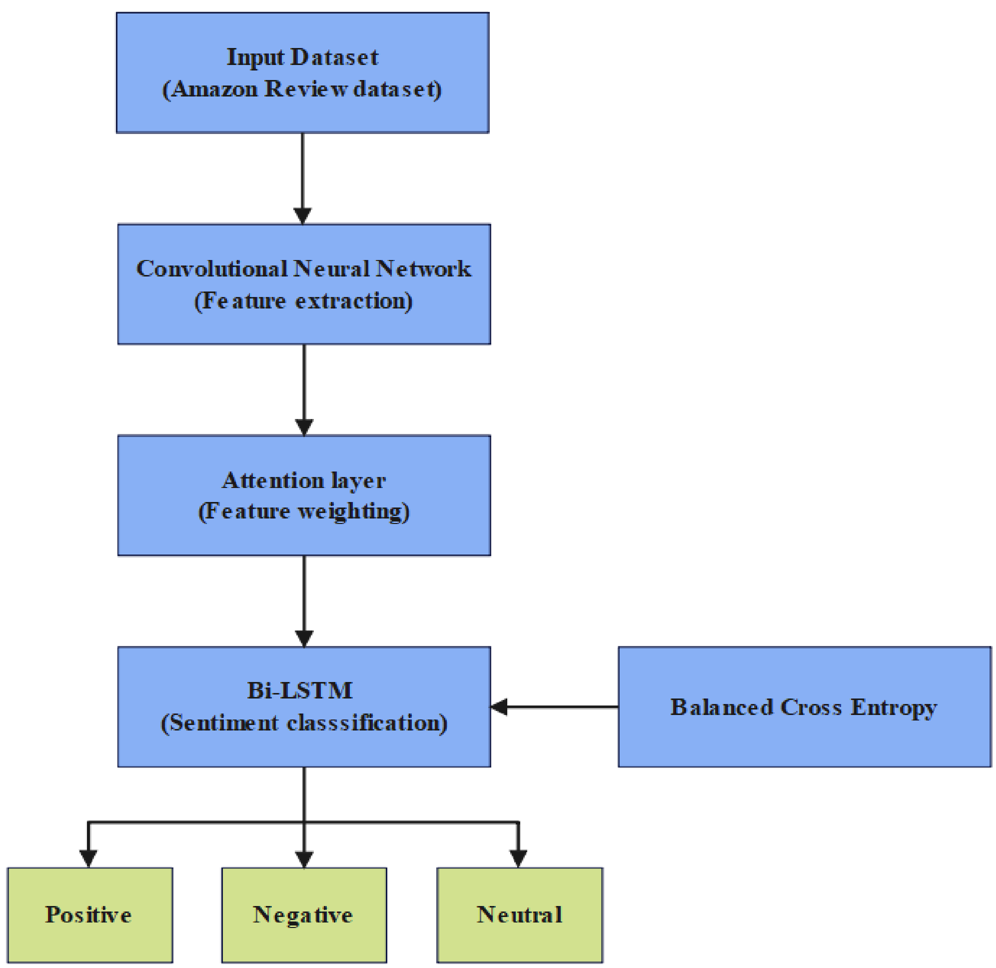

Fig. 4.1 :The flow of WEA-Bi-LSTM model for Sentiment Classification

4.4.1 Convolutional Neural Network

Humans find the process of understanding pictures intriguing, and this is a straightforward task for them. The machine that interprets images has many hidden intricacies. The CNN model is a deep learning system that seeks to mimic animal visual processing and is motivated by the visual cortex of the brain [86]. In the area of image processing, which includes detection, localization, segmentation, and classification, among other things, CNN's represents a giant step forward. The model is widely used primarily due to the great CNN efficiency in image categorization. The CNN model uses a convolutional layer with trainable weights and biases that are modelled after animal neurons. As demonstrated in Figure 4.2,Convolutional layers, fully linked layers, and activation functions are the core building elements of the CNN model. This paper provides a succinct explanation of the CNN concept, and the research includes a thorough description of CNN. [87, 88]

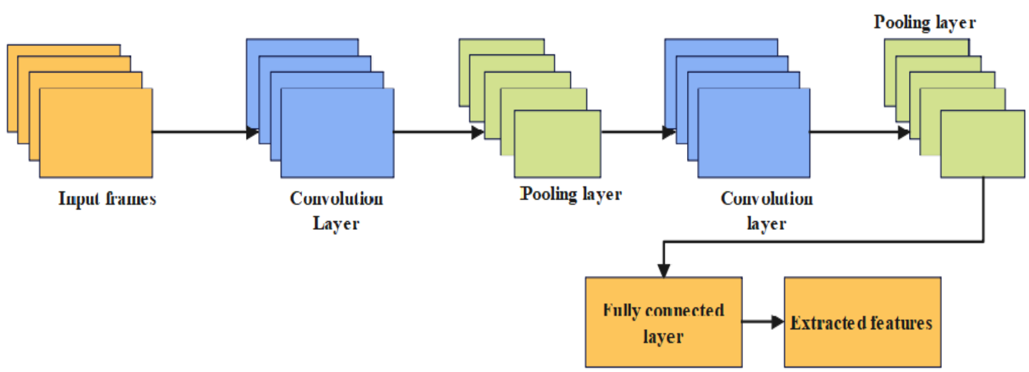

Fig. 4.2 :CNN Architecture model for Feature Extraction

4.4.1.1 Convolutional Layer:

The visual cortex's neuronal cells are involved in extracting characteristics from pictures in animal brains. Each neural cell extracts different characteristics that aid in the comprehension of a picture. Convolutional layers are used to simulate neuronal cells, and this allows for the extraction of properties including gradient direction, texture, colours, and edges. In convolutional layers and size, convolutional filter or kernels are learning filters is $n \times m \times d$, where image depth is d.During the forward pass, the Kernels are twisted over the input volume's height and width, andInput and filter entries are generated using the dot product. CNN learns which filters to use for texture, colour, edge, etc. An activation function layer is applied using the convolution layer's output.

4.4.1.2 Activation Function:

Since most real-world data is non-linear,Utilizing non-linear data transformation's activation functions. This guarantees that input space representation is translated to various output spaces in accordance with requirements.

This requires real-value number x and converting it into a range of 0 and 1. Large positive and negative inputs in particular are placed near 0 and unity, respectively. This is expressed in equation (4.1).

$$f(x) = \frac{1}{1+e^{-x}} \tag{4.1}$$

A number with real value x is taken into account in non-linear functions and converters x to 0 if x is negative. A common non-linear function that takes less time to compute and is better than the sigmoid is the ReLU activation function and $\tan h$ functions, as shown in Equation (4.2).

$$f(x) = \max(0, x) \tag{4.2}$$

4.4.1.3 Pooling

The pooling layer performs a non-linear down sampling's converged feature. As a result, dimensional reduction data processing requires less computer resources. To lower the spatial size, data are aggregated based on feature type or space; rotational variation in pictures overcomes translation; and overfitting is controlled. A rectangle patch set is created when inputs are partitioned using a pooling method. Based on the kind of pooling procedure, a single value is computed to replace each patch. Maximum pooling and average pooling are the two pooling types that are most often utilised.

4.4.1.4 Fully Connected Layer:

According to an artificial neural network, the inputs are linked to each node in the preceding layer, and weight values are given to each node this is known as a fully connected layer. The model's output is the total of the inputs multiplied by the appropriate weights. A completely linked layer is coupled to the sigmoid activation function to carry out the classification task. In Figure 4.3, can see CNN's completely linked layer.

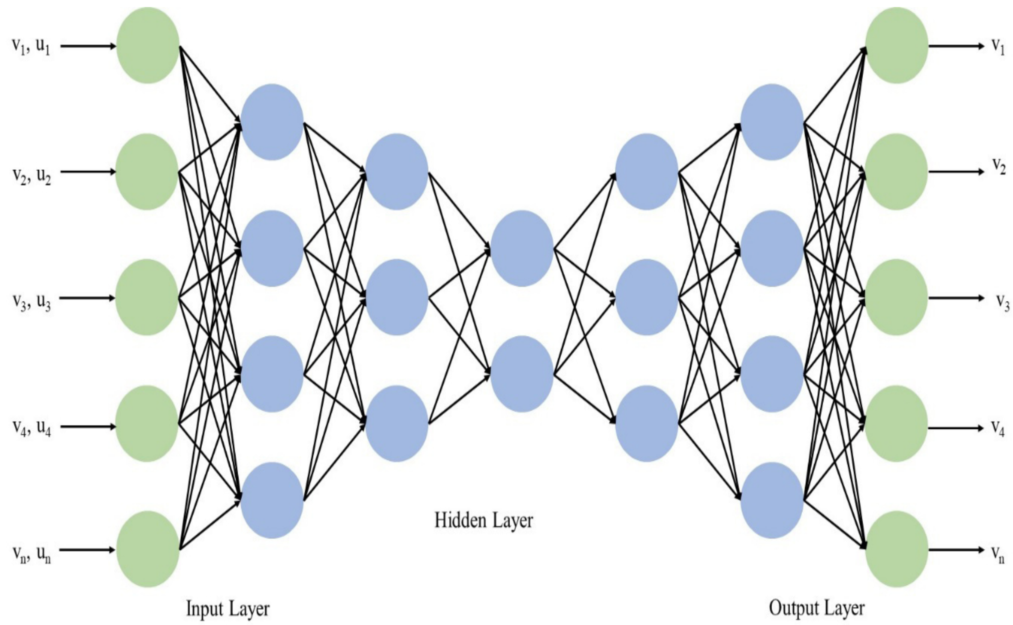

Fig. 4.3 :CNN model's fully Connected Layer

4.4.2 Attention Layer

Third region in the weighted word representation layer, CNN networks are used to utilise textual information. Each tweet is used as the basis for applying the word representation to a convolutional layer. The F_n Convolutional word vector matrix is computed using weighted matrix and is defined as $w \in R^{t \times m}$ a word vector is widely used to quantify local and fundamental features t is selected in F_n matrix, as in equation (4.3)

$$h_i = f(V_{i:j+t-1} \times W[i] + b_i) \quad (4.3)$$

Word vector when t design a feature map of $h \in R_{n-t+1}$, bias is b, The weight of the matrix is expressed as W, and non-linearity activation function is represented as f. Convolutional layers are used to create feature maps, while the max-pooling layer is utilized to lower the dataset's dimension and extract essential components, as indicated in equation (4.4).

$$p_i = Max[h_i] \quad (4.4)$$

where the feature map's max-pooling layer was applied to indicate $P_i \in R^{n-t+\frac{1}{2}}$.Maxpooling is used to extract equivalent essential components from the characteristics of different CNN layer filters without focusing on polarity or semantic relevance. An attention layer is used to give each feature on CNN-generated features value was shown in the equation (4.5).

$$A_i = \frac{\exp(p_i)}{\sum_i \exp(p_i)} \quad (4.5)$$

The above equation provides the max-pooling attention calculation and produces an attention score A_i is employed in every feature contextp_i.

To understand the feature context, the Bi-LSTM use the attention score output. Using a sequential map, the Bi-LSTM model creates the final features. The final feature context p_i is utilized for CNN's greatest feature map, and attention scores are shown as A_i.

4.4.3 Bi-LSTM Model

Due to its higher performance on sequences of data, the LSTM model is commonly employed in sentiment prediction. The input x_k at the current time step k and concealed layer result is denoted as h_{k-1} as earlier time step of concealed layer output. The LSTM cell state c_k and LSTM unit network architecture is shown in Figure 4.4.Since the gradient is propagated by the cell state, LSTM preserves the pertinent characteristics for a longer period of time than an RNN model.

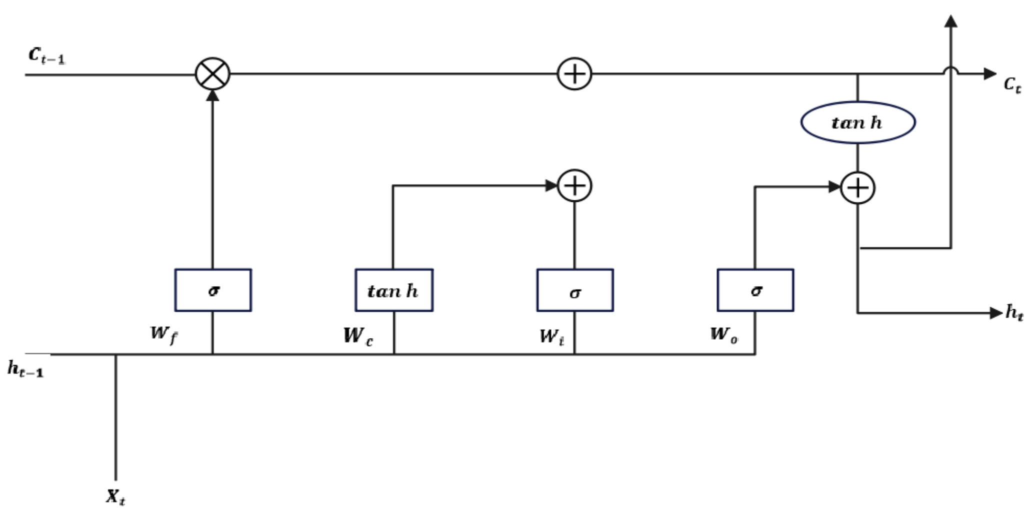

Fig. 4.4 :The LSTM unit Comprises of input, output and forget gate in Bi-LSTM

Forget gate f_k isFirst unit of the LSTM model, and the input gate is provided with information to update. Temporarily save the g_k new candidate value prior to updating a new cell state value. Input is received as $x_k$and h_{k-1},At each stage, the output is computed using biases, the weight parameter, and the ReLU activation function.. Equation (4.6-4.8) is used to measure gates and the ReLU activation function, which solves the vanishing gradient problem.

$$f_k = \sigma(W_x^f x_k + W_h^f h_{k-1} + b^f) \quad (4.6)$$

$$i_k = \sigma(W_x^i x_k + W_h^i h_{k-1} + b^i) \quad (4.7)$$

$$g_k = ReLU\left(W_x^g x_k + W_h^g h_{k-1} + b^g\right) (4.8)$$

The forget gate is used to apply the element-wise product of the cell state f_k at previous time step c_{k-1}, and new candidate value g_k isapplied to an input gate using the elementwise product i_k. For determining cell state, two determines sum c_k, as in equation (4.9).

$$c_k = f_k \odot c_{k-1} + i_k \odot g_k \quad (4.9)$$

The output gate o_k is last unit.This gate determines the output information Similar to the preceding procedures, the output gate is determined as shown in equation (4.10).

$$o_k = \sigma(W_x^o x_k + W_h^o h_{k-1} + b^o) \quad (4.10)$$

Equation (11) is used to compute the hidden state h_t.

$$h_k = o_k \odot ReLU(c_k) \quad (4.11)$$

Utilizing the LSTM model, which takes input from historical capacity data, the prediction is carried out $\{C_k\}$, size of input and output are same.Both the input and output sizes are the same, and the input and output data types are the same. Input and output formats are the foundation for the LSTM prediction.

Two order of each token in relation to its past and future context, two LSTM layers are used. The LSTM model analyses the left-to-right sequence and a separate sequence from right-to-left, as shown in Figure. 4.5. a front layer that is concealed $\vec{h}$ every time step, a hidden unit function t is depending on the input data step to calculate x_t and previous hidden state $\vec{h}_{t-1}$. The data entered at this stage x_t and the future hidden state $\overleftarrow{h}_{t+1}$ are to calculate the hidden unit function $\overleftarrow{h}$ in a hidden backward layer. The forward and backward context representations are used to produce $\vec{h}_t$and $\overleftarrow{h}_t$,that are joined together by a long vector, respectively. On combined results, forecasts' target signals are based.

The fully connected dense layer transforms the bi-directional network into a high-level sentiment representation to estimate text sentiment polarity. Equation (4.12) provide the output.

$$h_i = Relu(w_i h_p + b_i) \qquad (4.12)$$

The feature map that the Bi-LSTM network gives h_p, the obtained features is h_i, and learning new parameters are w_i, and b_i. The output layer's sentiment categorization is applied by the merge feature layer, as indicated in Figure 4.5.Application of sigmoid and sigmoid classifiers to binary and multi-class data is done in turn. Disparity between real and anticipated text sentiment is performed via balanced cross entropy.

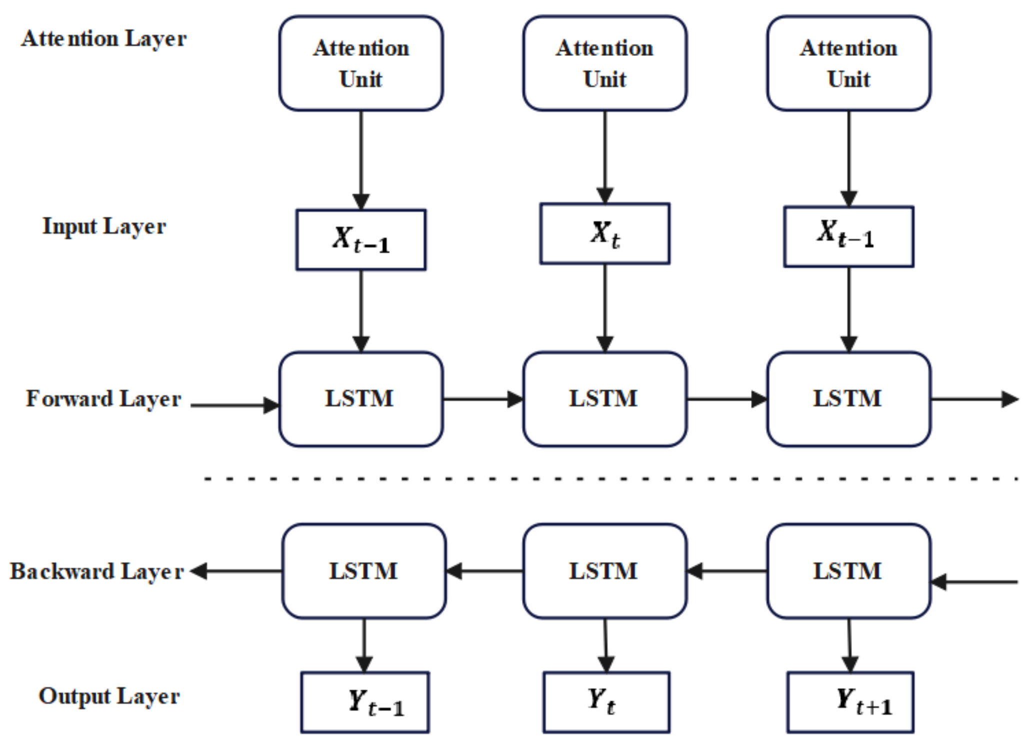

Fig. 4.5 :Attention layer in Bi-LSTM model for Sentiment Classification

4.4.4 Balanced Cross-Entropy

Two different kinds of optimization are paired with a tiny replay memory and the deep neural network baselinethe typical cross-entropy of softmax and distillation loss.

Total loss is used to train the model. L, to define weighted distillation loss L_d and average softmax cross-entropy loss L_c, as in equation (4.13)

$$L = \rho L_d + (1-\rho)L_c \qquad (4.13)$$

Where ρ is applied as N_{t-1}/N_t with incremental step t and $N_t$asamount of courses overall, to make up for two losses of significance.The amount of new classes utilized for learning is compared to the total number of classes used for learning at each training stage, and the relevance of distillation loss is changed.

Each time a new step is added t, the new parameters θ_t are utilized for initialization and the parameters from the previous phase. θ_{t-1}which, by use of distillation loss, is employed to preserve previously knowledge learned. The distillation loss L_d for each training sample $(x, y) \in X_t \cup X_M$, as in equation (4.14 & 4.15):

$$L_d(x) = \sum_{k=1}^{N_{t-1}} -\hat{p}_k(x)\log\big(p_k(x)\big)T^2, \qquad (4.14)$$

$$\hat{p}_k(x) = \frac{e^{\frac{\hat{z}_k(x)}{T}}}{\sum_{j=1}^{N_{t-1}} e^{\frac{\hat{z}_j(x)}{T}}}, \quad p_k(x) = \frac{e^{\frac{z_k(x)}{T}}}{\sum_{j=1}^{N_{t-1}} e^{\frac{z_j(x)}{T}}} \qquad (4.15)$$

Whenever y is the related ground truth label, the input data is x, the current model θ_t output logits is $z(x) = [z_1(x), \ldots, z_{N_t}(x)]$, and the previous incremental step θ_{t-1} output logits of model is $z(x) = [\hat{z}_1(x), \ldots, \hat{z}_{N_t} - 1\ (x)]$.

4.5 Results and Discussion

The Amazon dataset is used to apply the WEA-BiLSTM model for sentiment analysis, and it is contrasted with alternative approaches. For comparison, the WEA-BiLSTM model outputs were tested for performance and error metrics.

Table 4.1 :Performance of WEA-Bi-LSTM model for Various Iterations

Iteration	Precision (%)	Accuracy (%)	F-measure (%)	Recall (%)
0	0	0	0	0
50	81.4	91.7	76	72.3
100	82.1	92.5	76.5	73.4
150	82.3	92.8	77.2	73.6
200	82.5	93.4	77.8	73.8
250	83.3	93.7	78.1	74.1
300	84	94	78.3	75.2
350	85.7	94.9	78.7	76
400	86	95.6	80.4	76.1
450	86.8	97	80.4	76.3
500	86.8	97.4	81.2	76.7

The WEA-BiLSTM model is used for sentiment analysis, and Table 4.1 displays performance metrics for several iterations. This demonstrates that the model's ability to detect emotion gives it more accuracy. Word embedding is used in the attention technique's convolution process to improve feature representation. Because the model contains textual information on sarcasm and distinguishing features, their recall value is low. The WEA-BiLSTM model's accuracy is 95.6%, precision is 86%, recall is 76%, and F-Measure is 78.7% in the 400th iteration.

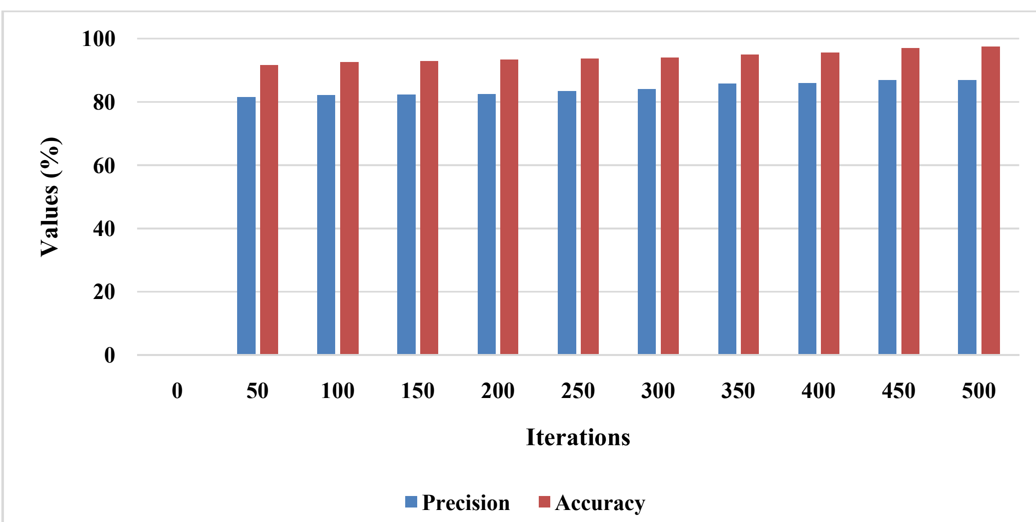

Fig. 4.6 :Accuracy and F-Measure of WEA-BiLSTM model for Various Iterations

Figure 4.6 displays the accuracy of the WEA-BiLSTM model and the F-measure for several rounds of sentiment analysis. This demonstrates that the WEA-BiLSTM model performs much worse than F-Measure while having better accuracy. The WEA-BiLSTM model performs less well in the negative class, which lowers the model's F-measure. Due to the user's caustic knowledge, the model performs significantly lower in the negative class.

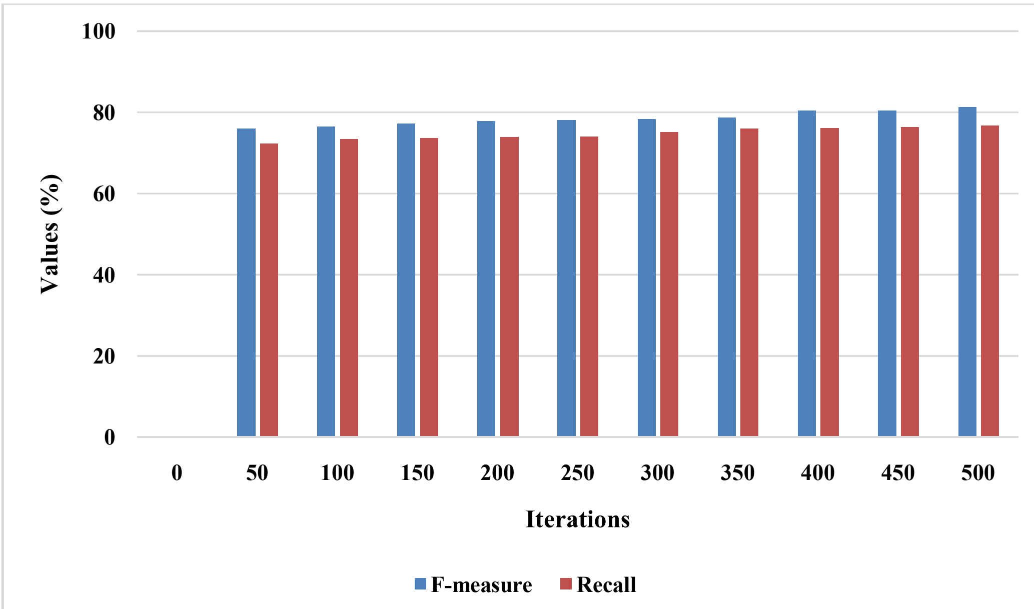

Fig. 4.7 :accuracy of the WEA-BiLSTM model

Figure 4.7 shows the accuracy and recall measurements for the WEA-BiLSTM model for several iterations. After 50 iterations, the WEA-BiLSTM model reaches the higher accuracy and recall levels. Based on the weight values supplied to the input text, the WEA approach accelerates the BiLSTM model's learning rate. The WEA approach gives terms with a strong connection to a class more weight, which boosts the model's effectiveness.

Table 4.2 :The WEA-BiLSTM model classifier comparison on Sentiment Analysis

Comparison	Precision (%)	Accuracy (%)	F-measure (%)	Recall (%)
SVM	81	91.8	75.6	73
ANN	71.9	81.9	76.8	72
LSTM	83	94.1	78.57	74.5
APSO-LSTM	85.28	96.8	80.04	76.08
CNN	85.6	97.1	80.64	76.33
Word Embedding CNN	86.8	97.4	81.2	76.7

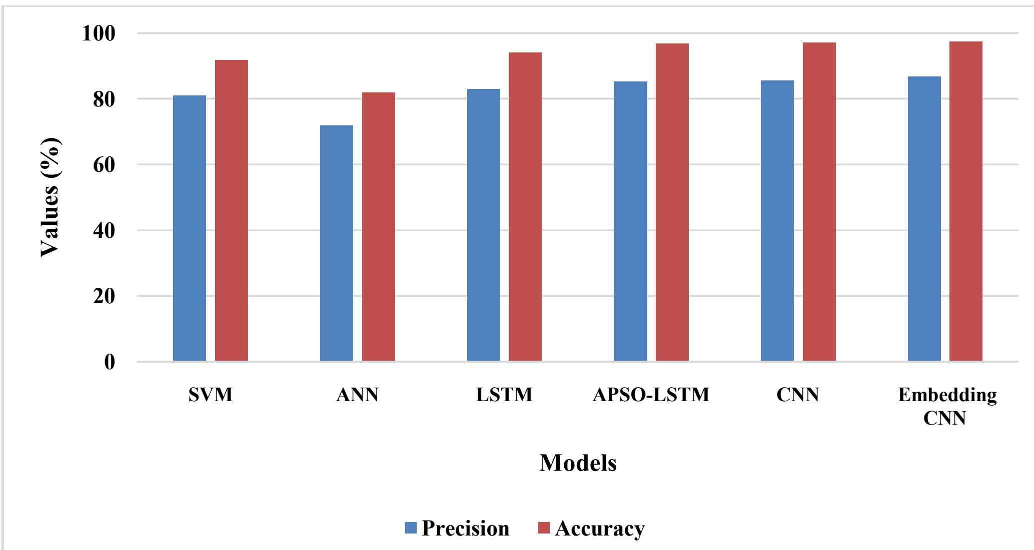

Fig. 4.8 :The WEA-BiLSTM model classifier comparison on Sentiment Analysis

Table 4.2 and Figure 4.8 compare the WEA-BiLSTM performance of the model to those of other existing classifiers. The WEA with CNN has the capacity to produce in-depth reports and directed outward with a strong connection to class greater weight. Existing SVM models have issues with imbalanced data, and LSTM models have issues with vanishing gradients. The APSO-LSTM factor represents certain ability to be flexible for classification and has a local optima trap. Due to the feature extraction procedure producing additional features, the CNN model suffers an overfitting issue. While the current CNN model has 97.1% accuracy and 85.4% precision in sentiment analysis, the WEA-BiLSTM model has 97.4% accuracy and 86.8% precision.

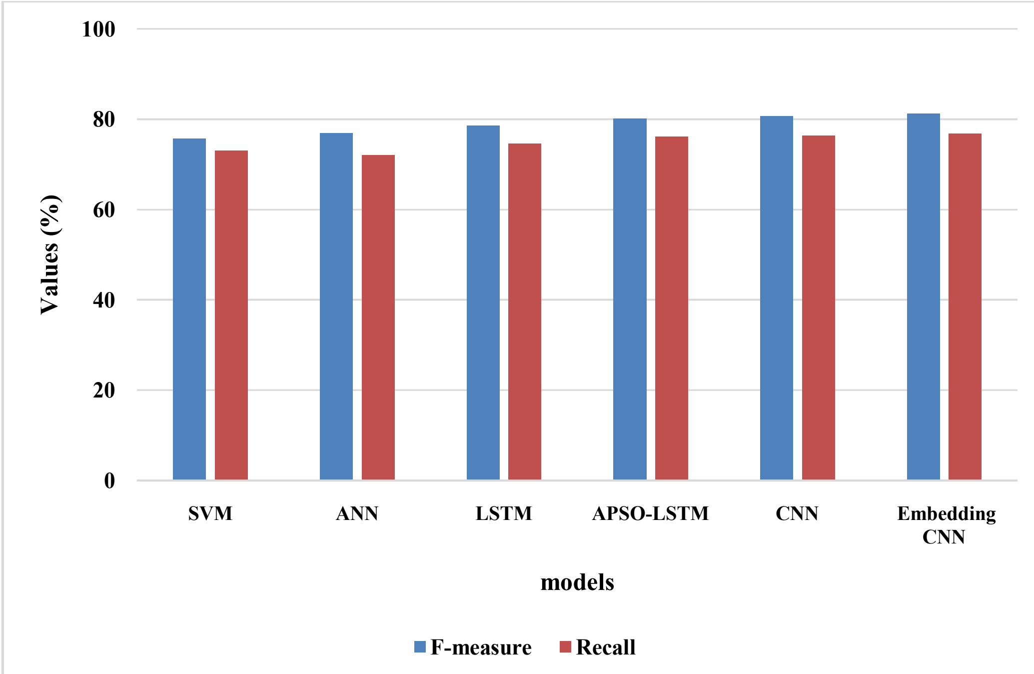

Fig.4.9 :WEA-BiLSTM precision and recall on Sentiment Analysis

Figure 4.9 shows the calculation of the WEA-BiLSTM model precision and recall value and a comparison with current sentiment analysis classifiers. The WEA-BiLSTM model employs embedding to provide phrases with strong relationships to classes a greater weight value. This improves the class-related performance of the WEA-BiLSTM model and raises its accuracy and recall values. Due to the addition of extra features in the convolutional layer, the existing CNN model performs less efficiently. The LSTM model's performance is decreased because to the vanishing gradient issue.

Table 4.3 :The WEA-BiLSTM model Performance for Various Categories

Category	Methods	AUC	Accuracy (%)	Recall (%)	F-measure (%)	Precision (%)
Books	IG	0.93	88	88	88	88
	CNN	0.94	90	90	90	90
	Proposed CNN	0.96	93	92	92	92
DVD	IG	0.94	90.25	90	90	90
	CNN	0.95	91.4	90	90	90
	Proposed CNN	0.97	93.5	93	93	93

Category	Methods	AUC	Accuracy (%)	Recall (%)	F-measure (%)	Precision (%)
Electronics	IG	0.95	91.63	92	92	92
	CNN	0.96	92.3	93	93	93
	Proposed CNN	0.97	93.7	95	95	95
Kitchen	IG	0.96	93.09	93	93	93
	CNN	0.97	94.1	93	93	93
	Proposed CNN	0.98	95.2	94	94	94
AVG	IG	0.94	90.74	90.75	90.75	90.75
	CNN	0.95	91.95	91.5	91.5	91.5
	Proposed CNN	0.97	93.85	93.5	93.5	93.5

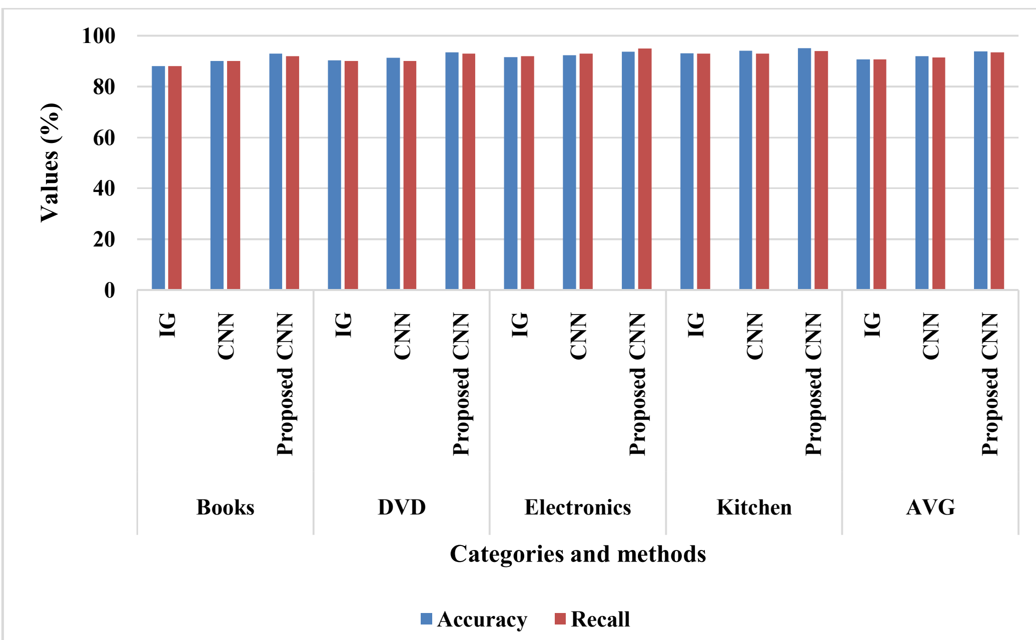

Fig. 4.10 :Accuracy and F-measure of the WEA-BiLSTM model for different Categories

According to Table 4.3 and Figure 4.10, the WEA-BiLSTM model's effectiveness is evaluated for a number of categories in the dataset. Every dataset category has a better performance for the WEA-BiLSTM model. The WEA-BiLSTM model uses word embedding method to provide terms with strong relationships to classes a greater weight value. The learning rate of the LSTM model in classifications is accelerated by the CNN features extracted and weight value. To solve the vanishing gradient problem, the balanced cross-entropy maintains the gradient values in the model. Both the WEA-BiLSTM and CNN models have accuracy and precision ratings of 93.85% and 93.5%, respectively.

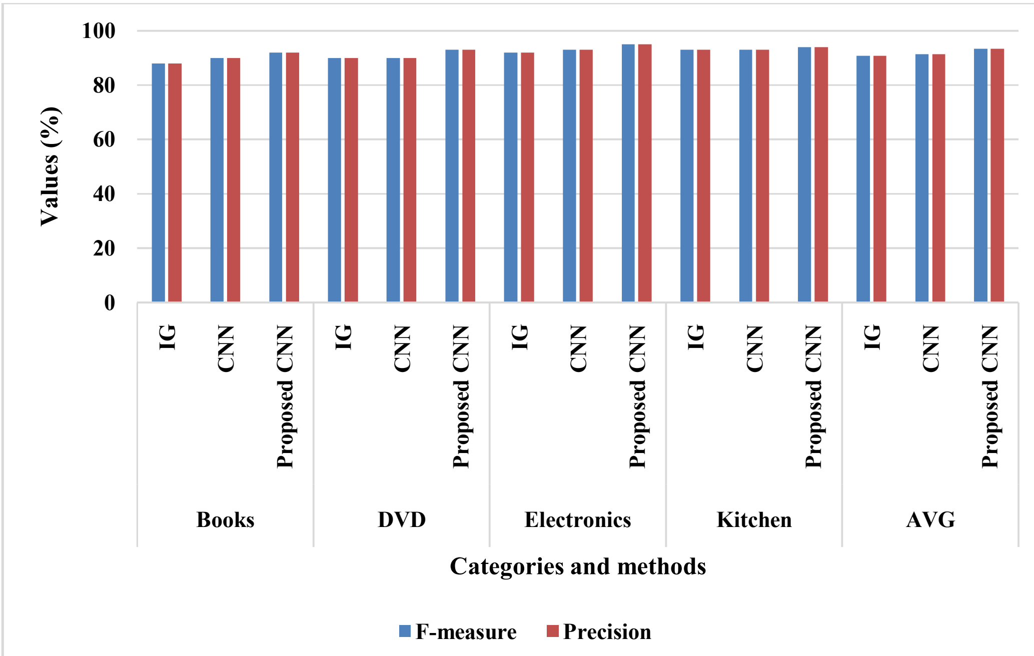

Fig. 4.11 :The WEA-BiLSTM model Precision and Recall for Various Categories

Figure 4.11 shows the accuracy and recall values for the WEA-BiLSTM model across the dataset's various categories. The WEA-BiLSTM model performs sentiment analysis with more precision and recall than other methods. The WEA-BiLSTM model gives terms with a strong connection to classes a greater weight value. This improves the model's performance class-wise and increases its accuracy and recall value.

Table 4.4 :WEA-BiLSTM model MAE on Sentiment Analysis

Datasets	ItemKNN	SlopeOne	A3NCF	SVD	CNN	UTER	Proposed CNN
Baby	0.8536	0.8557	0.8075	0.8123	0.7854	0.7966	0.7704
Sports and Outdoor	0.6771	0.663	0.633	0.6374	0.6087	0.6169	0.5843

Datasets	ItemKNN	SlopeOne	A3NCF	SVD	CNN	UTER	Proposed CNN
Pet Supplies	0.9202	0.8917	0.8746	0.88	0.8542	0.855	0.8412
Video Games	0.8482	0.8251	0.7916	0.7935	0.7891	0.7907	0.77

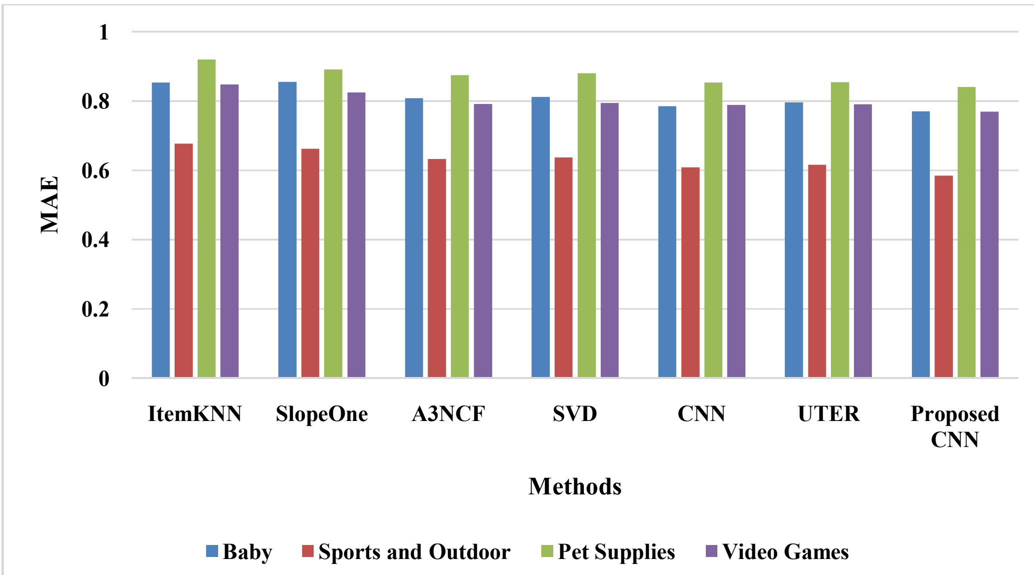

Fig. 4.12 :MAE of WEA-BiLSTM model on Sentiment Analysis

Table 4.4 and Figure 4.12 illustrate how the MAE value of the WEA-BiLSTM model for several categories on the Amazon dataset is determined and evaluated with modern techniques. The WEA-BiLSTM model performs better than the current approach. The WEA approach gives words with a strong connection to classes a greater weight value. The network's gradient value is sustained by balanced cross-entropy, which also eliminates the vanishing gradient problem.

Table 4.5 :WEA-BiLSTM RMSE on Sentiment Analysis

Datasets	ItemKNN	SlopeOne	A3NCF	SVD	CNN	UTER	Proposed CNN
Baby	1.1854	1.2012	1.0324	1.0704	0.9987	1.0146	0.9862
Sports and Outdoor	0.9959	1	0.8548	0.9065	0.8421	0.8501	0.8369

Datasets	ItemKNN	SlopeOne	A3NCF	SVD	CNN	UTER	Proposed CNN
Pet Supplies	1.2658	1.2454	1.1058	1.1448	0.9965	1.0627	0.9921
Video Games	1.1634	1.1421	1.0269	1.0511	0.9988	1.0164	0.9908

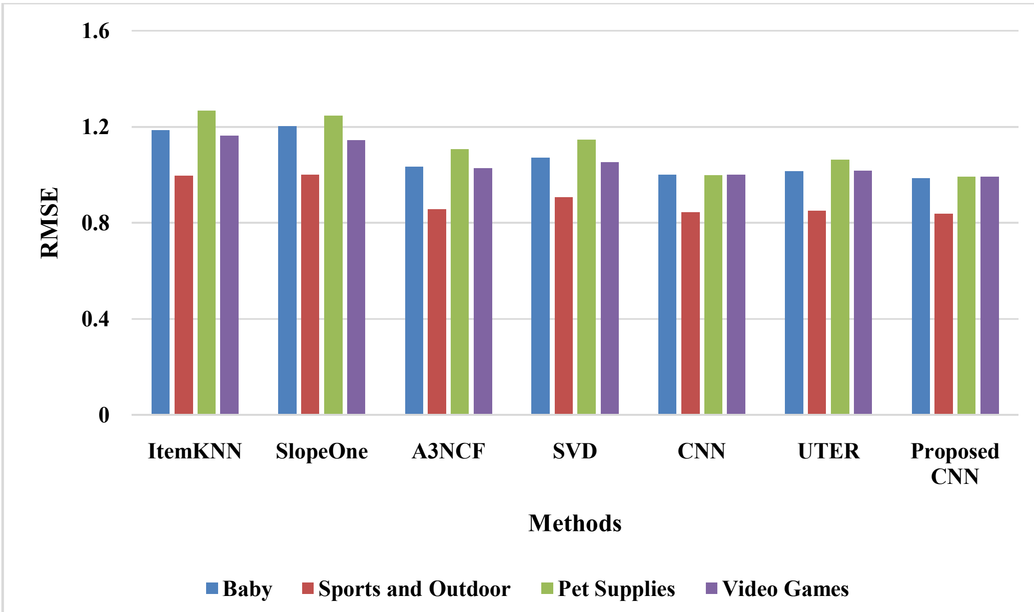

Fig. 4.13 :RMSE of WEA-BiLSTM model on sentiment analysis

Table 4.5 and Figure 4.13 illustrate how the WEA-BiLSTM model produces the RMSE value for several categories on the dataset. Due to the weight values that are provided in connection to classes, the WEA-BiLSTM model performs better in terms of classes. Due to the overfitting issue caused by the addition of extra features in convolutional layers, existing CNN-based models have limitations.

Table 4.6 :WEA-BiLSTM model for Sentiment Analysis

Algorithm	User Cold Start		Item Cold Start		User-Item Cold start	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
UTER	0.8123	1.2036	0.8106	1.2033	0.8112	1.2089
CNN	0.8108	1.1984	0.8065	1.1982	0.8057	1.2008
SVD	0.8409	1.2235	0.8446	1.2252	0.848	1.2276

Algorithm	User Cold Start		Item Cold Start		User-Item Cold start	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
A3NCF	0.8274	1.2137	0.8313	1.221	0.8336	1.2227
SlopeOne	0.8726	1.2626	0.8805	1.2639	0.8876	1.2683
ItemKNN	0.8653	1.251	0.8714	1.2501	0.8752	1.2532
Word Embedding CNN	0.8044	1.1920	0.7996	1.1905	0.8003	1.1941

Table 4.6 shows the MAE and RMSE values for the WEA-BiLSTM model evaluation for various cold start problems. Items with cold starts have less data about the products and less details about the users. The CNN-based feature extraction used by the WEA-BiLSTM model helps in extracting deep features and gives terms with strong relationships greater weight values.

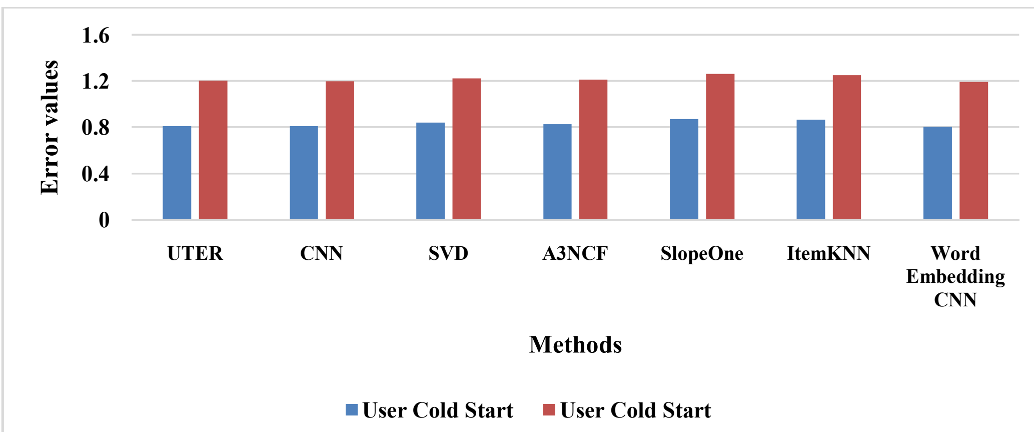

Fig.4.14 :The WEA-BiLSTM model for user cold start

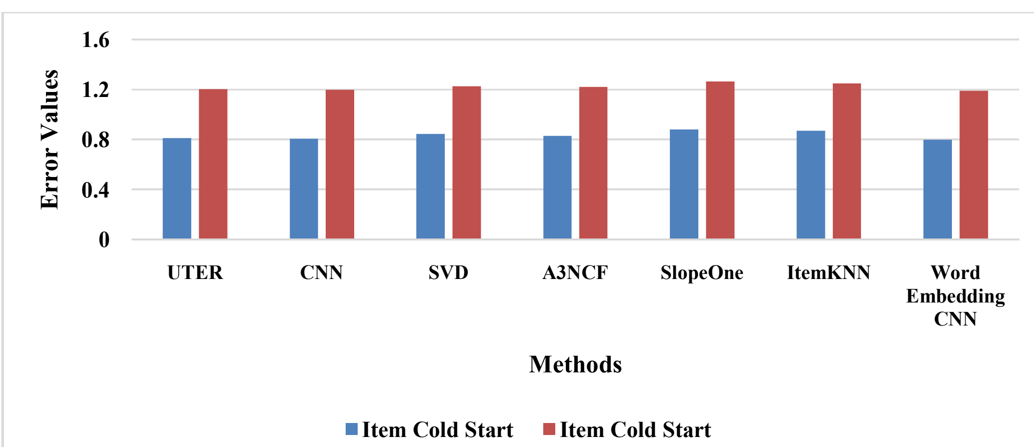

Fig. 4.15 :The WEA-BiLSTM model for Item cold start

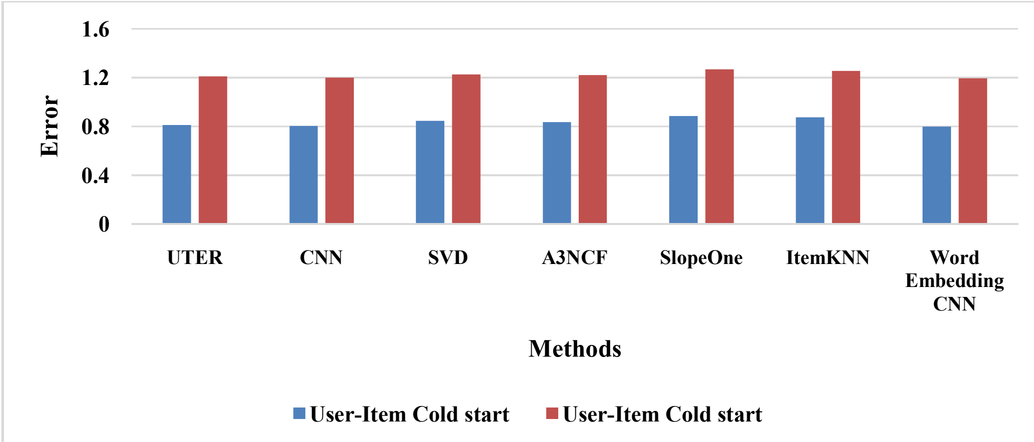

Fig. 4.16 :WEA-BiLSTM model for user Item cold start

According to Figures 4.14, 4.15, and 4.16, the MAE and RMSE metrics are used to analyze the WEA-BiLSTM model performance for user cold start, product cold start, and user-item cold start, respectively. The WEA-BiLSTM model performs sentiment analysis categorization more accurately than other models. Due to the model's utilization of CNN feature extraction for classification, the WEA-BiLSTM model performs better with lesser training data.

4.6 Summary

Sentiment is the process of identifying the intended sentiments that are conveyed in text-based materials like social media postings, product reviews, and online discussion environments. Similar to how sentiments derived from customer reviews and comments are often applied by businesses to develop and improve their goods and postal service. The user's sentiment about the product is provided through sentiment analysis on reviews, which is important for product development. The vanishing gradient problem and the overfitting issue in classification were constraints of the existing LSTM and CNN-based models used in sentiment analysis. This study suggested the WEA approach to give words with a strong relationship to classes a greater weight value. Due to the convolutional layer generating more features, the CNN model is used for feature extraction. To maintain the gradient in the system and solve the vanishing gradientproblem, balanced cross entropy is used. With less training data required for classification, improved performance is achieved by CNN feature extraction. Future work on this methodology will focus on using the feature selection strategy to choose pertinent features and improve performance in imbalance data.

CHAPTER – V

CONCLUSION AND FUTURE SCOPE

With the quickerenhancement of e-commerce systems in the past couple of years, product review sentiment analysis technology has gained increasing attention. Because the quantity of users searching the internet is increasing exponentially, the developer network is increasingly concentrating on the user expertise of searching. A customer evaluated if the product is terrible or excellent from web sites. It is very crucial for any company to be aware of client feedback on any given item. Sentiment analysis on review offers information about product's user sentiment, which is beneficial for product development. Conventional LSTM and CNN-based models with vanishing gradient as well as overfitting problems in classification were used in sentiment analysis.

In the first study, proposed a method for predicting the effective online shopping sites, a customer survey determined whether the product is poor or excellent from enterpriseonline portals. Prioritizing is of utmost importance for any business to be aware of customer feedback on any particular item. The Amazon datasets will be used in this research to analyse and classifypreviousevaluators on e-commerce platforms and their impact on itemprominence.This study proposes to review the posting process and develop a deep learning method for reviewer projection. Dependent on the decisionassessment, a suggestedviewpointspatial analysis is used to sort the basicvalues. By taking into account the pair of the repetition of viewpoints and the effect of decisions, yielding the suggestedonline portals. In the second study, proposed the WEA technique to provide higher weight values to the words having strong relation with classes. The CNN model is applied for feature extraction due to generation of more features in the convolutional layer. The balanced cross entropy is applied to maintain the gradient in the network to solves vanishing gradient problem. The CNN feature extraction helps to provide higher performance for less number of training data for classification.

5.1 Future Scope

- In the future, an accurate system will be implemented to enhance the classification accuracy for recommendation of product.
- In future, novel feature selection technique will be analyzed to select relevant features and achieve higher performance in imbalance data.

- Instead of supervised MLP classifier approach, an optimal cost unsupervised clustering techniques will be deployed to identify the span reviews hidden in the review dataset.
- A refined and definite ranking algorithm could be evolved in future to provide optimistic product ranking for predicting the opinion orientation score in the review sentences.

REFERENCES

References

1. Moorthi, K., GauravDhiman, P. Arulprakash, C. Suresh, and K. Srihari. "A survey on impact of data analytics techniques in E-commerce." Materials Today: Proceedings (2021).
2. Yi, Shanshan, and Xiaofang Liu. "Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review." Complex & Intelligent Systems 6, no. 3 (2020): 621-634.
3. Kurdi, B., Muhammad Alshurideh, ImanAkour, H. Alzoubi, Bader Obeidat, and Ahmad Alhamad. "The role of digital marketing channels on consumer buying decisions through eWOM in the Jordanian markets." *International Journal of Data and Network Science* 6, no. 4 (2022): 1175-1186.
4. Wu, Yuanyuan, Eric WT Ngai, Pengkun Wu, and Chong Wu. "Fake online reviews: Literature review, synthesis, and directions for future research." *Decision Support Systems* 132 (2020): 113280.
5. Padwal, Nilam Deepak, SampadaGulavani, and Rajesh Kanthe. "Application of Opinion Mining and Sentiment Analysis." (2022).
6. Shaukat, Zeeshan, Abdul AhadZulfiqar, Chuangbai Xiao, Muhammad Azeem, and Tariq Mahmood. "Sentiment analysis on IMDB using lexicon and neural networks." *SN Applied Sciences* 2, no. 2 (2020): 1-10.
7. Yadav, Rohan Kumar, Lei Jiao, Morten Goodwin, and Ole-ChristofferGranmo. "Positionless aspect based sentiment analysis using attention mechanism." *Knowledge-Based Systems* 226 (2021): 107136.
8. Edalati, Maryam, Ali Shariq Imran, ZenunKastrati, and Sher Muhammad Daudpota. "The potential of machine learning algorithms for sentiment classification of students' feedback on MOOC." In *Proceedings of SAI Intelligent Systems* pp. 11-22. Springer, Cham, 2022.
9. Madasu, Avinash, and SivasankarElango. "Efficient feature selection techniques for sentiment analysis." *Multimedia Tools and Applications* 79, no. 9 (2020): 6313-6335.
10. Ayetiran, Eniafe Festus. "Attention-based aspect sentiment classification using enhanced learning through CNN-BiLSTM networks." *Knowledge-Based Systems* 252 (2022): 109409.
11. Álvarez-Carmona, Miguel Á., Ramón Aranda, Rafael Guerrero-Rodríguez, Ansel Y. Rodríguez-González, and A. Pastor López-Monroy. "A combination

of sentiment analysis systems for the study of online travel reviews: Many heads are better than one." *Computación y Sistemas* 26, no. 2 (2022).

12. Elmitwally, NouhSabri. "Building A Multi-class XGBoost Model for Arabic Figurative Language." *International journal on Computer and Information Sciences*, pp. 1-4. IEEE, 2020.
13. Kumar, Avinash, Savita Sharma, and Dinesh Singh. "Sentiment Analysis on Twitter Data using a Hybrid Approach‖." *International conference of Computer Sciences and Engineering* 7.
14. Deniz, Ayça, MerihAngin, and PelinAngin. "Sentiment and Context-refined Word Embeddings for Sentiment Analysis." *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 927-932. IEEE, 2021.
15. Ray, Paramita, and AmlanChakrabarti. "A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis." *Applied Computing and Informatics* (2020).
16. Du, Qiming, Nan Li, Wenfu Liu, Daozhu Sun, Shudan Yang, and FengYue. "A Topic Recognition Method of News Text Based on Word Embedding Enhancement." *Computational Intelligence and Neuroscience* 2022 (2022).
17. Zucco, Chiara, Barbara Calabrese, Giuseppe Agapito, Pietro H. Guzzi, and Mario Cannataro. "Sentiment analysis for mining texts and social networks data: Methods and tools." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10, no. 1 (2020): e1333.
18. Wankhade, Mayur, Annavarapu Chandra SekharaRao, and ChaitanyaKulkarni. "A survey on sentiment analysis methods, applications, and challenges." *Artificial Intelligence Review* (2022): 1-50.
19. Mowlaei, Mohammad Erfan, Mohammad SanieeAbadeh, and HamidrezaKeshavarz. "Aspect-based sentiment analysis using adaptive aspect-based lexicons." *Expert Systems with Applications* 148 (2020): 113234.
20. Sharma, ShashankShekhar, and GautamDutta. "SentiDraw: Using star ratings of reviews to develop domain specific sentiment lexicon for polarity determination." *Information Processing & Management* 58, no. 1 (2021): 102412.
21. Rice, Douglas R., and Christopher Zorn. "Corpus-based dictionaries for sentiment analysis of specialized vocabularies." *Political Science Research and Methods* 9, no. 1 (2021): 20-35.

22. Grljević, Olivera, ZitaBošnjak, and AleksandarKovačević. "Opinion mining in higher education: a corpus-based approach." *Enterprise Information Systems* 16, no. 5 (2022): 1773542

23. Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN Computer Science* 2, no. 3 (2021): 1-21.

24. Mitra, Ayushi. "Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset)." *Journal of Ubiquitous Computing and Communication Technologies (UCCT)* 2, no. 03 (2020): 145-152.

25. Karimova, Farida. "A survey of e-commerce recommender systems." *European Scientific Journal* 12, no. 34: 75-89.

26. Yang, L., Li, Y., Wang, J. and Sherratt, R.S., 2020. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*, *8*, pp.23522-23530.

27. Li, L., Goh, T.T. and Jin, D., 2020. How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. *Neural Computing and Applications*, *32*(9), pp.4387-4415.

28. Basiri, M.E., Abdar, M., Cifci, M.A., Nemati, S. and Acharya, U.R., 2020. A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. *Knowledge-Based Systems*, *198*, p.105949.

29. Li, W., Zhu, L., Shi, Y., Guo, K. and Cambria, E., 2020. User reviews: Sentiment analysis using lexicon integrated two-channel CNN–LSTM family models. *Applied Soft Computing*, *94*, p.106435.

30. Sahoo, S.R. and Gupta, B.B., 2021. Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, *100*, p.106983.

31. Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R. and Mora, H., 2020. A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, *90*, pp.523-537.

32. Islam, M.R., Liu, S., Wang, X. and Xu, G., 2020. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining*, *10*(1), pp.1-20.

33. Rehman, A.U., Malik, A.K., Raza, B. and Ali, W., 2019. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications*, *78*(18), pp.26597-26613.

34. Wang, W.M., Wang, J.W., Li, Z., Tian, Z.G. and Tsui, E., 2019. Multiple affective attribute classification of online customer product reviews: A heuristic deep learning method for supporting Kansei engineering. *Engineering Applications of Artificial Intelligence*, *85*, pp.33-45.

35. Sharma, S.K., Chakraborti, S. and Jha, T., 2019. Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach. *Information Systems and e-Business Management*, *17*(2), pp.261-284.

36. Vermeer, S.A., Araujo, T., Bernritter, S.F. and van Noort, G., 2019. Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, *36*(3), pp.492-508.

37. Arora, M. and Kansal, V., 2019. Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis. *Social Network Analysis and Mining*, *9*(1), pp.1-14.

38. Banbhrani, S.K., Xu, B., Lin, H. and Sajnani, D.K., 2022. Spider Taylor-ChOA: Optimized Deep Learning Based Sentiment Classification for Review Rating Prediction. *Applied Sciences*, *12*(7), p.3211.

39. Chen, J., Yu, J., Zhao, S. and Zhang, Y., 2021. User's Review Habits Enhanced Hierarchical Neural Network for Document-Level Sentiment Classification. *Neural Processing Letters*, *53*(3), pp.2095-2111.

40. Wei, S. and Song, S., 2022. Sentiment Classification of Tourism Reviews Based on Visual and Textual Multifeature Fusion. *Wireless Communications and Mobile Computing*, *2022*.

41. Geetha, M.P. and Renuka, D.K., 2021. Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*, *2*, pp.64-69.

42. Priyadarshini, I. and Cotton, C., 2021. A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. *The Journal of Supercomputing*, *77*(12), pp.13911-13932.

43. Shobana, J. and Murali, M., 2021. An efficient sentiment analysis methodology based on long short-term memory networks. *Complex & Intelligent Systems*, *7*(5), pp.2485-2501.

44. Zhang, Y., Liu, Z. and Sang, C., 2021. Unifying paragraph embeddings and neural collaborative filtering for hybrid recommendation. *Applied Soft Computing*, *106*, p.107345.

45. Onan, A., 2021. Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and Computation: Practice and Experience*, *33*(23), p.e5909.

46. Zhao, H., Liu, Z., Yao, X. and Yang, Q., 2021. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. *Information Processing & Management*, *58*(5), p.102656.

47. Khan, A., Gul, M.A., Uddin, M.I., Ali Shah, S.A., Ahmad, S., Firdausi, A., Dzulqarnain, M. and Zaindin, M., 2020. Summarizing online movie reviews: a machine learning approach to big data analytics. *Scientific Programming*, *2020*.

48. Chatterjee, S., Goyal, D., Prakash, A. and Sharma, J., 2021. Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application. *Journal of Business Research*, *131*, pp.815-825.

49. Yi, S. and Liu, X., 2020. Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. *Complex & Intelligent Systems*, *6*(3), pp.621-634.

50. Noori, B., 2021. Classification of customer reviews using machine learning algorithms. *Applied Artificial Intelligence*, *35*(8), pp.567-588.

51. Arasu, B.S., Seelan, B.J.B. and Thamaraiselvan, N., 2020. A machine learning-based approach to enhancing social media marketing. *Computers & Electrical Engineering*, *86*, p.106723.

52. Dadhich, A. and Thankachan, B., 2021. Social & Juristic challenges of AI for Opinion Mining Approaches on Amazon &Flipkart Product Reviews Using Machine Learning Algorithms. *SN Computer Science*, *2*(3), pp.1-21.

53. Li, L., Rong, S., Wang, R. and Yu, S., 2021. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and

process control in drinking water treatment: A review. *Chemical Engineering Journal*, *405*, p.126673.

54. Abo, M.E.M., Idris, N., Mahmud, R., Qazi, A., Hashem, I.A.T., Maitama, J.Z., Naseem, U., Khan, S.K. and Yang, S., 2021. A multi-criteria approach for arabic dialect sentiment analysis for online reviews: Exploiting optimal machine learning algorithm selection. *Sustainability*, *13*(18), p.10018.
55. Nilashi, M., Ahmadi, H., Arji, G., Alsalem, K.O., Samad, S., Ghabban, F., Alzahrani, A.O., Ahani, A. and Alarood, A.A., 2021. Big social data and customer decision making in vegetarian restaurants: A combined machine learning method. *Journal of Retailing and Consumer Services*, *62*, p.102630.
56. Soumya, S. and Pramod, K.V., 2020. Sentiment analysis of malayalam tweets using machine learning techniques. *ICT Express*, *6*(4), pp.300-305.
57. Ahmad, A., Feng, C., Khan, M., Khan, A., Ullah, A., Nazir, S. and Tahir, A., 2020. A systematic literature review on using machine learning algorithms for software requirements identification on stack overflow. *Security and Communication Networks*, *2020*.
58. Al-Garadi, M.A., Hussain, M.R., Khan, N., Murtaza, G., Nweke, H.F., Ali, I., Mujtaba, G., Chiroma, H., Khattak, H.A. and Gani, A., 2019. Predicting cyberbullying on social media in the big data era using machine learning algorithms: review of literature and open challenges. *IEEE Access*, *7*, pp.70701-70718.
59. Martinez-Torres, M.D.R. and Toral, S.L., 2019. A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, *75*, pp.393-403.
60. Gamal, D., Alfonse, M., M. El-Horbaty, E.S. and M. Salem, A.B., 2018. Analysis of machine learning algorithms for opinion mining in different domains. *Machine Learning and Knowledge Extraction*, *1*(1), pp.224-234.
61. Portugal, I., Alencar, P. and Cowan, D., 2018. The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, *97*, pp.205-227.
62. Rathor, A.S., Agarwal, A. and Dimri, P., 2018. Comparative study of machine learning approaches for Amazon reviews. *Procedia computer science*, *132*, pp.1552-1561.

63. Ghosh, M. and Sanyal, G., 2018. An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning. *Journal of Big Data*, *5*(1), pp.1-25.

64. Huh, J.H., 2018. Big data analysis for personalized health activities: machine learning processing for automatic keyword extraction approach. *Symmetry*, *10*(4), p.93.

65. Chen, Z., Van Khoa, L.D., Teoh, E.N., Nazir, A., Karuppiah, E.K. and Lam, K.S., 2018. Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge and Information Systems*, *57*(2), pp.245-285.

66. Zhao, Zhijie, Jiaying Wang, Huadong Sun, Yang Liu, Zhipeng Fan, and FuhuaXuan. "What factors influence online product sales? Online reviews, review system curation, online promotional marketing and seller guarantees analysis." *IEEE Access* 8 (2019): 3920-3931.

67. Bi, Jian-Wu, Yang Liu, Zhi-Ping Fan, and Erik Cambria. "Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model." *International Journal of Production Research* 57, no. 22 (2019): 7068-7088.

68. Singh, JyotiPrakash, SedaIrani, Nripendra P. Rana, Yogesh K. Dwivedi, Sunil Saumya, and Pradeep Kumar Roy. "Predicting the “helpfulness” of online consumer reviews." *Journal of Business Research* 70 (2017): 346-355.

69. Qi, Jiayin, Zhenping Zhang, SeongminJeon, and Yanquan Zhou. "Mining customer requirements from online reviews: A product improvement perspective." *Information & Management* 53, no. 8 (2016): 951-963.

70. Wang, Yuren, Xin Lu, and Yuejin Tan. "Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines." *Electronic Commerce Research and Applications* 29 (2018): 1-11.

71. Yang, Qinjuan, YanghuiRao, HaoranXie, Jiahai Wang, Fu Lee Wang, and Wai Hong Chan. "Segment-level joint topic-sentiment model for online review analysis." *IEEE Intelligent Systems* 34, no. 1 (2019): 43-50.

72. Dang, Yan, Yulei Zhang, and Hsinchun Chen. "A lexicon-enhanced method for sentiment classification: An experiment on online product reviews." *IEEE Intelligent Systems* 25, no. 4 (2009): 46-53.

73. Farhadloo, Mohsen, Raymond A. Patterson, and Erik Rolland. "Modeling customer satisfaction from unstructured data using a Bayesian approach." *Decision Support Systems* 90 (2016): 1-11.

74. Nandal, Neha, RohitTanwar, and JyotiPruthi. "Machine learning based aspect level sentiment analysis for Amazon products." *Spatial Information Research* 28, no. 5 (2020): 601-607.

75. Zhao, Huiliang, Zhenghong Liu, Xuemei Yao, and Qin Yang. "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach." *Information Processing & Management* 58, no. 5 (2021): 102656.

76. Bilal, M. and Almazroi, A.A., 2022. Effectiveness of Fine-Tuned BERT Model in Classification of Helpful and Unhelpful Online Customer Reviews. *Electronic Commerce Research*, pp.1-21.

77. Suresh, P. and Gurumoorthy, K., 2022. Mining Of Customer Review Feedback Using Sentiment Analysis For Smart Phone Product. In *International Conference on Computing, Communication, Electrical and Biomedical Systems* (pp. 247-259). Springer, Cham.

78. Subba, B. and Kumari, S., 2022. A heterogeneous stacking ensemble based sentiment analysis framework using multiple word embeddings. *Computational Intelligence*, *38*(2), pp.530-559.

79. Deniz, E., Erbay, H. and Coşar, M., 2022. Multi-label classification of e-commerce customer reviews via machine learning. *Axioms*, *11*(9), p.436.

80. Al Kilani, N., Tailakh, R. and Hanani, A., 2019, October. Automatic classification of apps reviews for requirement engineering: Exploring the customers need from healthcare applications. In *2019 sixth international conference on social networks analysis, management and security (SNAMS)* (pp. 541-548). IEEE.

81. Chen, D., Zhang, D. and Liu, A., 2019. Intelligent Kano classification of product features based on customer reviews. *CIRP Annals*, *68*(1), pp.149-152.

82. Noori, B., 2021. Classification of customer reviews using machine learning algorithms. *Applied Artificial Intelligence*, *35*(8), pp.567-588.

83. Khan, M. and Malik, K., 2019. Sentiment classification of customer’s reviews about automobiles in roman urdu. In *Future of Information and Communication Conference* (pp. 630-640). Springer, Cham.

84. Sun, X., Han, M. and Feng, J., 2019. Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products. *Decision Support Systems*, *124*, p.113099.

85. Geetha, M., Singha, P. and Sinha, S., 2017. Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. *Tourism Management*, *61*, pp.43-54.

86. Chen, W., Jiang, M., Zhang, W.G. and Chen, Z., 2021. A novel graph convolutional feature based convolutional neural network for stock trend prediction. Information Sciences, 556, pp.67-94

87. Nguyen, H.D., Tran, K.P., Thomassey, S. and Hamad, M., 2021. Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. International Journal of Information Management, 57, p.102282

88. Cao, J., Li, Z. and Li, J., 2019. Financial time series forecasting model based on CEEMDAN and LSTM. Physica A: Statistical mechanics and its applications, 519, pp.127-139.

PUBLICATIONS

DISSEMINATION OF RESEARCH WORK

1. V Sagvekar, Prashant Sharma, "Study on Product Opinion Analysis for Customer Satisfaction on E-Commerce Websites", published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).doi:10.3233/APC210206, © 2021 The authors and IOS Press.

2. Presented paper on "Product opinion analysis for customer satisfaction on E-commerce websites", in international virtual conference on 15-16th January 2022 held at Pacific Academy of Higher Education and Research University, Udaipur.

3. Patent published on 10/2/2023 Application No.202311005260 A .Title of the invention: Predicting Overall Customer Satisfaction For An Effective Product Opinion Analysis On E-Commerce Websites.

4. Sagvekar, V. R., & Sharma, P. (2023). Word embedding attention and balanced cross entropy technique for sentiment analysis. Multiagent and Grid Systems, 19(1), 23-42.

5. Paper title," Weighted Ensemble LSTM Model with Word Embedding Attention for E-commerce Product Recommendation", submitted and is under review.

Recent Trends in Intensive Computing
M. Rajesh et al. (Eds.)

doi:10.3233/APC210206

Study on Product Opinion Analysis for Customer Satisfaction on E-Commerce Websites

Ms. Vijaya Sagvekar [a, 1], Dr. Prashant Sharma [b]

[a] *Research Scholar, Department of Computer Engineering, PAHER, Udaipur, India*

[b] *Associate Professor & HOD Computer Science & Engineering Department PAHER, Udaipur (Rajasthan),* India

Abstract---- The E-commerce websites have been emerged in a high range of marketing benefits for the users to publish or share the experience of the received product by posting review that contain useful comments, opinions and feedback on the product. These days, a large number of clients acquire freedoms to look at comparative items in online sites and pick their top choices in computerized retailers, like Amazon.com and Taobao.com. Client audits in online media and electronic trade Websites contain important electronic word data of items. Sentiment Analysis is broadly applied as voice of clients for applications that target showcasing and client care. Sentiment extractors in their most essential structure classify messages as either having a good or negative or once in a while neutral supposition. A typical application of sentiment investigation is the programmed assurance of whether an online review contains a positive or negative review. Subsequently, in this paper, with the use of the strategies on sentiment analysis, obstinate sentences alluding to a particular element are first recognized from item online audits. We have proposed deep learning strategy as a classification model for discovering the condition of review. The outcomes showed suggested site for the client dependent on the early audits, past reviews and answer given to inquiry audit for the client. Additionally, it is seen that the proposed strategy can ready to answer every one of the reviews with a superior closeness like a human reaction to the client.

Keywords--- E-commerce, opinion analysis, sentiment analysis, deep learning

1. Introduction

The marketers and manufacturers have been focused on the market performance from long time. The development of product marketing strategy for managing the product quality is helpful in making a better decision shows improvement in the performance [1]. The manufacturers and advertisers continuously gather the product information that useful in analyzing the performance in the market. In the traditional methods, the main data sources for analysis of main data sources were collected from the manufacturers‟ internal data, offline customer reviews by the surveys, review forms should be in handwritten format etc [2].

[1] Ms. Vijaya Sagvekar, Research Scholar, Department of Computer Engineering, PAHER, Udaipur
Email: vijayasagvekar@gmail.com.

As the technology became advanced, the e-commerce websites enabled to publish the opinion of users on product and the users got a public platform to share their useful comments and opinions towards the product they purchased. According to the survey, 97 % of consumers will be influenced for purchasing the product after going through the reviews posted by other customers. So, a most of clients will see online reviews prior to settling purchase decision [3,4].

91 % of people regularly or occasionally read online reviews before purchasing a product and also the early reviews on product before purchasing has high impact on succeeding the product sales [5]. Even though, the early reviewers contribute small proportion of reviews, it is easy to determine the failure or success of new services and new products. Based on the early reviews from the early reviewers helps to adjust marketing strategies in product improvement on designs and helps in succeeding the new product [6]. Based on the early reviews, the companies will recognize the early reviews thereby improve the product designs, marketing strategies that lead to a success on newly launched products. Thus, early analysts become the accentuation to screen and pull in at the early advancement phase of an organization in the world [7-9]. The early reviews have pulled in showcasing experts broadly investigate the customer buy goals. For example one of the largest e-commerce companies is Amazon in the world, where it provides early reviewer program opportunities for making early reviews that helps the company to acquire those early reviews on products that have few reviews or no reviews. With the help of Amazon shoppers program, will provide information about buying a product by making smarter decision. Based on the above discussions, we can see that for product marketing the most important is early reviewers of product [10-14].The present research will take initiative for studying the behavior characteristics for the early reviewers and posting it in e-commerce platform such as Amazon, Yelp. The main aim of the research work is to analyze the consumer satisfaction on products bought from E-commerce sites and improved the performance. The overall characteristics of early reviews have to be analyzed from the early reviewers needed to be compared to majority and laggard reviewers. An early reviewer tends to post more helpful reviews and helps to gives a higher average rating score to products. The rating behaviours are characterized that helps to find the scores received from others that helps to determine the correlation of the reviews based on the product popularity [15]. The discoveries with the character factors are connected with the hypothesis as follows: higher normal rating scores can be considered as the ideal attitude towards the items, and higher support votes of early surveys given by others can be seen as an intermediary proportion of the opinion leadership.

2. Literature Review

Many researches have been developed for predicting the ranking based on the online reviews. Some of the studies are as follows. In the literature survey, a survey of recent techniques is highlighted with its advantage and limitations.

Table 1. Literature Review

Authors	Method	Advantage	Disadvantage
Ahani, A., et al [16]	Self-Organizing Map and Higher Order Singular Value decomposition clustering algorithm	The hybrid algorithm was used for assisting that overcame the data related complications for online reviews and presented spa hotel market segmentation for predicting the travel choice using machine learning algorithms.	The available customer data from TripAdvisor included only the general preferences of spa hotel customers degraded the performance rate

Rita, P et al [17]	Four-dimensions of e-service quality model	The four dimensions of e-service quality were considered such as the impact on customer trust, satisfaction, customer behaviour and building existed literature based on e-service quality during online shopping.	The quality of online stores, in general, was not based on the product segments sold in online stores and the measurements were not applicable for assessing product segments.
Lucini, F.R., et al [18]	Texting Mining approach Latent Dirichlet Allocation (LDA)	The developed model presented a novel framework for customer satisfaction measuring in the airline industry using Latent Dirichlet Allocation (LDA) that detected the popular topic using the natural language processing and machine learning process	The data restricted the diversity of opinions as the proficient in English were in more and likely provided airline experience information.
Bai, T. et al [19]	Margin based Embedding Ranking Model (MERM)	The characterized and predicted the early reviewers for E-commerce sites to present effective product marketing. The developed model used Margin based Embedding Ranking Model (MERM) that predicted the early reviewers in a cold-start setting	The developed model used Margin based Embedding Ranking Model (MERM) that predicted the early reviewers in a cold-start setting.
Zhao, Y.,et al [20]	Technical attributes and sentiment polarity	The review samples were taken from trip advisor that predicted overall customer satisfaction using technical attributes in online and textual reviews of customers.	The textual reviews are influenced by languages and are different cultures needed extension for examining different language reviews.
Liu, Y. et al [21]	Product Competitive and Quality Management and Marketing Strategy	The developed a product competitive advantage analysis for providing an essential basis for quality management and marketing strategy development on social media. The novel method provided an essential basis for managing the marketing strategy and quality using user generated content	The customer comments relied which comes after target product availability of the customers and was not applicable during the design and development of the product.
Jian Jin et al [22]	Product feature extraction and sentiment analysis	The developed model performed opinionated representative for specific product based on the features especially for the competitive products. The sentimental analysis was performed for opinionated sentences that refer specific features for online reviews.	During choosing comparative sentences in the review that were of many for different products resulted lower information comparativeness values.
Sun, Q., et al [23]	Sentiment analysis eWOM	The developed model extracted large volume of online customer reviews and performed sentiment analysis for eWOM products. The developed model used semi-supervised fuzzy product ontology mining algorithm for extraction of features with negative or positive labels.	The developed model required improvement in positive and negative opinion words extraction and also polarity computation.
Thompson, J.J., et al [24]	Lexicon-Based Sentiment Extraction Analysis	The developed model performed sentimental analysis for chat messaging by the player who was involved in video game StarCraft 2. The developed model performed sentimental analysis was applicable for toxicity detection and also identified the players and their messages that are threat for the player.	The human raters disagree not only the sentiment but also disagreed the toxicity that suggested that the task is not straight forward and thus the performance at high rate was impossible

Kumar, S., et al [25]	Sentimental Analysis For The Product Review For EEG Response	The developed model performed multimodal framework for the estimation of product rating on customer product and their brands. The reviews obtained from the global viewers were processed using Natural Language Processing (NLP) technique that computed the score for global rating.	The emotional state needed to be considered for improving the performance.

3. Motivation for Study

There are numerous quantities of studies that examine the monetary results or the drivers of online reviews and, simultaneously, propose suggestions for the plan of review frameworks, for example, giving reviewers a predefined review format. Albeit a significant number of review framework configuration highlights have been proposed throughout the long term, truth be told, not very many have really been dissected. The new online plans of action and conditions have arisen including two-sided stage organizations (e.g., Uber). These empower two-sided reviews and require adjusted plan highlights to, for example, alleviate correspondence in two-sided review frameworks. Finally, most plan highlights of review frameworks have been investigated for fixed gadgets like PCs. Notwithstanding, online reviews are progressively delivered and burned through cell phones, which require explicit plan highlights. These are the unsolved problems in the ranking analysis for online reviews and many research work based on this problem are still examining.

4. Motivation for Study

There are numerous quantities of studies that examine the monetary results or the drivers of online reviews and, simultaneously, propose suggestions for the plan of review frameworks, for example, giving reviewers a predefined review format. Albeit a significant number of review framework configuration highlights have been proposed throughout the long term, truth be told, not very many have really been dissected. The new online plans of action and conditions have arisen including two-sided stage organizations (e.g., Uber). These empower two-sided reviews and require adjusted plan highlights to, for example, alleviate correspondence in two-sided review frameworks. Finally, most plan highlights of review frameworks have been investigated for fixed gadgets like PCs. Notwithstanding, online reviews are progressively delivered and burned through cell phones, which require explicit plan highlights. These are the unsolved problems in the ranking analysis for online reviews and many research work based on this problem are still examining.

5. Proposed Methodology

In our paper, the proposed a application for predict online items in different E-commerce sites like Amazon, Paytm, Flipkart, ShopClues and Snapdeal in reviews of clients. Firstly, we processed the datasets of different items of particular organization from E-commerce sites. Each product has its own features for categorized depends on a particular feature. This features such as significantly positively or negatively influences customer ratings, readability, subjectivity, length–significantly, and sentiment polarity. The entire surveys of client that refreshed itemsbased are generated for collecting FCM model [27].

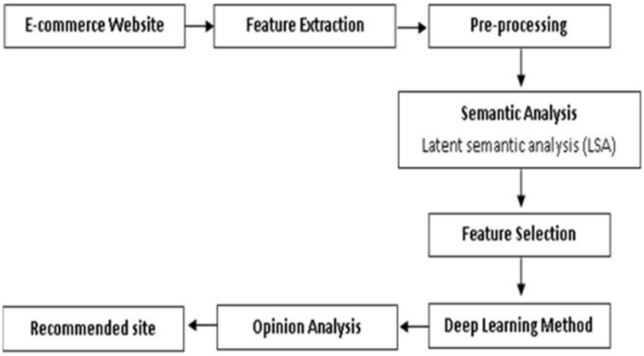

Figure 1. Flow graph of Proposed Method

Dataset
The proposed research uses a row dataset obtained from online E-shopping sites (ShopClues, Paytm, Flipkart, Amazon and Snapdeal). Each and every site is important in introducing a chance to buy the product by great population. The popularity of these sites is identified by various administrations that help to improve the distribution of various products belonging to various categories. Among the E-commerce database, Amazon dataset is used for the present research that has around142 million product reviews taken for the duration May 1996 to July 2014. Each of the review from the site consisted of textual comment that was posted by the product user that accompanied to publish time stamp accurate in the study. Usually the reviews are associated scale that is having up to five star of rating that is associated with textual description. A review an audit is related with a rating score in a five-star scale. Every item is related with a category label and a textual description.

Feature Extraction The reviews from the dataset are extracted using feature extraction technique that reviews of collected item from the E-commerce sites improved the quality of review during analysis. The feature extraction includes the items enhancement for performing audit rating of each item. The attributes here refers to the cost, quality, positive survey. The feature extraction mines the opinion of customer reviews, summarize the reviews, store and produce step preparation.
Pre-processing Transforming text into something an algorithm can process is a confounded process. In this part, the means associated with text processing are as per the following. (i) Tokenization: The superfluous tokens are not difficult to filter, where a document is changed into passages or sentences into words. In our work, the online review is tokenizing into words. (ii) Removal of Unnecessary labels and Punctuation: The following stage is to eliminate Punctuation, as the Punctuation doesn't do any additional data while treating text data. (iii) Removing stop words: Frequent words such as “the”, “is”, etc. that do not have explicit semantic. (iv) Lemmatization: Another way to deal with eliminate affectation by deciding the grammatical form and using point by point data set of the language.
Semantic Analysis The semantic analysis is performed for the preprocessed data where, the natural language content reads all the words captures the content by performing real meaning of any text. The text elements are assigned and are assigned based on the logical and grammatical role. The semantic analysis analyzes the surrounding context in the text to accurately disambiguate the exact meaning of words. The relationship

between the concepts in the text is also developed to identify the most relevant elements in text and understand the topic discussed. In the semantic analysis, Latent Semantic Analysis (LSA) was used in NLP that analyses the relationship among the set of documents and the terms they contain by creating a set of concepts related to the documents and terms.

Feature Selection From the extracted concepts, feature selection process is performed for selecting the subset terms that were occurred in training and these selected subsets were treated as features that performed text classification. Firstly, the training is performed and is applied for the classifier decrease the size of the vocabulary effectively. Secondly, feature selection improves the classification accuracy thereby eliminates the noise features in figure1.

Deep Learning the deep learning methods provide an opportunity that faces challenges in NLP problems such as sequence-to-sequence prediction. The developed model performed deep learning methods for learning the features based on the NL which is required by the model specifies the required features and were extracted. In natural language processing, the performance of deep learning is depending on genuine outcomes and that the enhancements give off an impression of being proceeding and maybe speeding up.

Opinion Analysis The optimal ranking opinion assesses the significance of each element relatively with respect to the sentiment score that utilized for rank highlights measuring. Based on the opinion analysis, the commentators rated the sites that are having vital data distinguished the untruthful opinions.

Recommended Site Based on these recommendations, a perspective positioning calculation were performed for ranking the vital angles resulted a viewpoint recurrence that impacted opinions for each perspective for general sentiments.

6. Results and Discussion

For accurate recommendation of products a new opinion analysis system is developed, by analyzing the reviews that users are posted for the products. The main aim of this research is to create a accurate keyword extraction technique and clustering approach for recommending the products using amazon customer review dataset interms of positive and negative form. In this paper, with GWO algorithm, a keyword extraction method (LDA) is used for selecting the proper key words. The acquired same keywords are clustered using PFCM algorithm. The developed recommendation system has main advantage is that, system has ability to find the fake products, keep track of clients satisfaction etc.

Table 2. Comparison between proposed systems and different classifiers.

	Precision	Recall	F1-Measure	AUC	Accuracy
Random Forest	69.003	73.25	71.132	48.263	73.187
Decision Tree	75.766	75.583	74.618	54.628	75.528
Proposed	76.106	76.781	74.167	57.404	77.236

The proposed system conveyed a powerful execution through quantitative analysis and comparative analysis. From the test analysis, the proposed system accomplished around 77.236% 77.236% of classification accuracy, but the existing techniques achieved limited accuracy in amazon customer review dataset. In future work, an accurate system is implemented for further improvement in the classification accuracy for recommendation of product.

7. Conclusion

Now a days the developer community is more and more focusing on the user experience of browsing, because number of users browsing the internet are exponential increased. In this paper, proposed the approach to predict the most effective web-based shopping sites. By client survey determined whether poor or great is the product from business websites. It is more important for any business to have the knowledge about the reviews of customer regarding its any particular items.

In this paper, to analyze the characteristics and categorize early reviewers on an e-shopping sites and their effect on product popularity will use the Amazon datasets. This proposed study review posting process and develop a deep learning model for the prediction of reviewers. We also mine the summary of surveys, customer reviews. Based on the opinion analysis, a recommended perspective positioning computation to rank the fundamental points by contemplating both the viewpoint recurrence and the impact of opinions gives the recommended site.

References

[1] Smith, A.N., Fischer, E. and Yongjian, C., 2012. How does brandrelated user-generated content differ across YouTube, Facebook, and Twitter?. Journal of interactive marketing, 26(2), pp.102-113.

[2] Dong, J., Li, H. and Zhang, X., 2014, November. Classification of customer satisfaction attributes: an application of online hotel review analysis. In Conference on e-Business, e-Services and e-Society (pp. 238-250). Springer, Berlin, Heidelberg.

[3] De Giovanni, P. and Zaccour, G., 2014. A two-period game of a closed-loop supply chain. European Journal of Operational Research, 232(1), pp.22-40.

[4] Farhadloo, M., Patterson, R.A. and Rolland, E., 2016. Modeling customer satisfaction from unstructured data using a Bayesian approach. Decision Support Systems, 90, pp.1-11.

[5] Elish, M.O., 2014. A comparative study of fault density prediction in aspect-oriented systems using MLP, RBF, KNN, RT, DENFIS and SVR models. Artificial Intelligence Review, 42(4), pp.695-703.

[6] Clavel, C. and Callejas, Z., 2015. Sentiment analysis: from opinion mining to human-agent interaction. IEEE Transactions on affective computing, 7(1), pp.74-93.

[7] Nahm, Y.E., 2013. A novel approach to prioritize customer requirements in QFD based on customer satisfaction function for customer-oriented product design. Journal of Mechanical Science and Technology, 27(12), pp.3765-3777. Yang, L., Geng, X. and Liao, H., 2016. A web sentiment analysis method on fuzzy clustering for mobile social media users. EURASIP Journal on Wireless Communications and Networking, 2016(1), p.128.

[8] Yang, J.B., Wang, Y.M., Xu, D.L., Chin, K.S. and Chatton, L., 2012. Belief rule-based methodology for mapping consumer preferences and setting product targets. Expert Systems with Applications, 39(5), pp.4749-4759.

[9] Decker, R. and Trusov, M., 2010. Estimating aggregate consumer preferences from online product reviews. International Journal of Research in Marketing, 27(4), pp.293-307. [11] Dehkharghani, R., Mercan, H., Javeed, A. and Saygin, Y., 2014. Sentimental causal rule discovery from Twitter. Expert Systems with Applications, 41(10), pp.4950-4958.

[10] Chen, A., Lu, Y. and Wang, B., 2017. Customers" purchase decisionmaking process in social commerce: a social learning perspective. International Journal of Information Management, 37(6), pp.627-638.

[11] Uysal, A.K. and Gunal, S., 2014. The impact of preprocessing on text classification. Information Processing & Management, 50(1), pp.104- 112.

[12] Zhang, W., Yoshida, T. and Tang, X., 2011. A comparative study of TF* IDF, LSI and multi-words for text classification. Expert Systems with Applications, 38(3), pp.2758-2765.
[13] Zhao, T. and Eskenazi, M., 2016. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning. arXiv preprint arXiv:1606.02560.
[14] Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L. and Weaven, S., 2019. Market segmentation and travel choice prediction in Spa hotels through TripAdvisor‟s online reviews. International Journal of Hospitality Management, 80, pp.52-77.
[15] Rita, P., Oliveira, T. and Farisa, A., 2019. The impact of e-service quality and customer satisfaction on customer behavior in online shopping. Heliyon, 5(10), p.e02690.
[16] Lucini, F.R., Tonetto, L.M., Fogliatto, F.S. and Anzanello, M.J., 2020. Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. Journal of Air Transport Management, 83, p.101760.
[17] Bai, T., Zhao, W.X., He, Y., Nie, J.Y. and Wen, J.R., 2018. Characterizing and predicting early reviewers for effective product marketing on e-commerce websites. IEEE Transactions on Knowledge and Data Engineering, 30(12), pp.2271-2284.
[18] Zhao, Y., Xu, X. and Wang, M., 2019. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. International Journal of Hospitality Management, 76, pp.111-121. [
[19] Liu, Y., Jiang, C. and Zhao, H., 2019. Assessing product competitive advantages from the perspective of customers by mining usergenerated content on social media. Decision Support Systems, 123, p.113079.
[20] Jin, J., Ji, P. and Gu, R., 2016. Identifying comparative customer requirements from product online reviews for competitor analysis. Engineering Applications of Artificial Intelligence, 49, pp.61-73.
[21] Sun, Q., Niu, J., Yao, Z. and Yan, H., 2019. Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. Engineering Applications of Artificial Intelligence, 81, pp.68-78.
[22] Thompson, J.J., Leung, B.H., Blair, M.R. and Taboada, M., 2017. Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model. Knowledge-Based Systems, 137, pp.149-162. [25] Kumar, S., Yadava, M. and Roy, P.P., 2019. Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. Information Fusion, 52, pp.41-52.
[23] Pradip Suresh Mane, Ashok Kumar Jetawat, Pravin Jagannath Nikumbh‟, 2020, "Web Page Recommendation using Random Forest with Fire Fly Algorithm in Web Mining", International Journal of Engineering and Advanced Technology (IJEAT), Vol. 9, Issue-3, February 2020, pp. 499-505.
[24] Kolhe L.N., Khairnar V., Jetawat A.K. :Prediction-Based Parallel Clustering Algorithm for M-Commerce. In: Fong S., Akashe S., Mahalle P. (eds) Information and Communication Technology for Competitive Strategies. Lecture Notes in Networks and Systems, vol 40. Springer, Singapore, pp. 31-39

Multiagent and Grid Systems – An International Journal 19 (2023) 23–42
DOI 10.3233/MGS-221511
IOS Press

Word embedding attention and balanced cross entropy technique for sentiment analysis

Vijaya Ravindra Sagvekar[a,*] and Prashant Sharma[b]
[a]*Department of Computer Engineering, Pacific Institute of Technology, Rajasthan, India*
[b]*Department of Computer Science and Engineering, Pacific University, Udaipur, Rajasthan, India*

Received 2 November 2022
Accepted 29 March 2023

Abstract. Customer feedback is useful for product development and increases the sales of the product. Reviews on e-commerce websites provided by the user provide valuable information about the product. Sentiment analysis on the text review helps to analyze the sentiment of users about the product and predict the sales of a product. The existing techniques in sentiment analysis use Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models that have limitations of vanishing gradient problem and overfitting problem. Initially, the Amazon review dataset is collected and processed in word embedding stage. The CNN is utilized to extract the features from the input dataset for sentiment analysis. The Word Embedding Attention (WEA) technique provides higher weight to the words having strong relation with class. The CNN feature helps to provide higher performance for a smaller number of training data. This technique helps to increase the performance of the model related class-wise, thus increasing the precision and recall value. Finally, WEA technique in Bi-directional LSTM (BiLSTM) is used to increase the classification performance. The Balanced Cross-Entropy is proposed to maintain the gradient and solves the vanishing gradient problem in the network. The WEA-BiLSTM model has 97.4% accuracy, and 86.8% precision, and the existing CNN model has 97.1% accuracy and 85.4% precision in sentiment analysis. In this study, WEA-LSTM is used for the sentimental analysis of user reviews. This technique solves the vanishing gradient problem in the network by using Balanced Cross Entropy and helps to increase the performance of the model.

Keywords: Balanced cross-entropy, bi-directional long short-term memory, convolutional neural network, sentiment analysis, word embedding attention

1. Introduction

Consumer feedback is useful information in business to assess the quality and improve the product for the benefit of customers that provides an idea of expectation of new products. Various deep learning techniques were applied to accurately predict the customer's opinion on mobile phone reviews of Amazon products [1]. On e-commerce websites such as Flipkart, eBay, and Amazon, thousands of users leave reviews about the products and services provided by the websites such as Yelp, Rotten Tomatoes, and Trip Advisor. Some users leave reviews about the product or services on social media. Therefore, customer

*Corresponding author: Vijaya Ravindra Sagvekar, Department of Computer Engineering, Pacific Institute of Technology, Rajasthan, India. E-mails: vijayasagvekar@outlook.com and vijayasagvekar@gmail.com.

ISSN 1574-1702/$35.00

reviews and feedback shared on the common platform about services or products influence new customers' perspectives towards institutions, organizations, services, and products [2]. In natural language processing, one of the basic tasks is to learn low-dimensional word vector representation from a large dataset. The existing word embedding techniques learns word vectors from context semantic information and grammar while ignoring words sentiment information. Some techniques of sentiment information in reviews don't consider certain words in various domains [3]. Sentiment analysis exploits the grouping of text mining, computational linguistics, and natural language processing to evaluate, calibrate, derive and analyze textual data in terms of documents, phrases, sentences, etc. Recently, natural language processing gained a huge consideration due to its efficiency in extracting useful information [4]. Sentiment analysis aims to classify the given text into three categories of user emotions: Positive, Neutral, and Negative [5].

Sentiment analysis is applied to input text to determine the sentiment polarity such as positive, neutral, and negative. Recently, online merchants and retailers ask their customers to share their opinion about the product. Consequently, more opinions are created every day and this is hard to process the data to extract useful information about the customer review [6]. Recently, deep learning technologies were applied to sentiment analysis to provide more promising results. Various neural network-based architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) were applied to improve the efficiency of sentiment analysis [7]. Deep learning models of Convolutional Neural Network (CNN) and LSTM provide higher efficiency in sentiment analysis [8–10]. The objectives and contribution of this study are given below:

1) The word embedding technique is proposed to provide higher weight value to the words having strong relation to the classes. This helps to increase the learning rate of the model and increases the class-wise performance of the model.
2) The Balanced Cross Entropy is applied to maintain the gradient in the network to reduce the vanishing gradient problem. The gradient is maintained in the network based on the loss calculation.
3) The CNN-based feature extraction technique is applied to generate more features in the convolutional layer and helps the model to provide higher performance for less training data.
4) The Word Embedding Attention-Bi-directional Long Short-Term Memory (WEA-BiLSTM) model has higher performance in sentiment analysis classification compared to existing techniques. The WEA-BiLSTM model solves the vanishing gradient problem and decreases the overfitting problem in the classification.

This research study is structured as follows: Sentiment analysis researches are reviewed in Section 2 and the Word Embedding Attention (WEA) and Balanced Cross Entropy explanation are given in Section 3, the results of WEA-BiLSTM model are specified in Section 4. Lastly, Section 5 provides the conclusion.

2. Literature review

Sentiment analysis helps to find the response of the user related to the product and this technique can be used for sales prediction or to improve the product quality. Sentiment analysis is a challenging process in large e-commerce review data due to the presence of unstructured data. Recent research related to sentiment analysis was reviewed in this Section.

2.1. Feature extraction methods

Banbhrani et al. [11] applied the sentiment analysis technique for the prediction of review ratings based on feature extraction techniques. The model provides significant features such as number of sentences, emoticons, hashtags, elongated words, punctuation marks, numerical words, number of capitalized words, Term Frequency Inverse Document Frequency (TF-IDF), and SentiWordNet-based statistical features. Sentiment classification was performed using random multimodal deep learning and extracted features. The developed technique was a supervised technique and deep learning technique has an overfitting problem.

Wei and Song [12] construct a classification model based on text sentiment and a multi-input matrix was used to combine sentiment features that provide input of multi-channel CNN to extract sentiment features for text sentiment classification. The model also developed an image sentiment classification model using face images and global image merging. The supervision module on CNN with weighted loss was applied to extract features of facial sentiment, whole image sentiment was fused with facial target sentiment to determine image sentiment polarity. The text and image sentiment output were fused for the decision fusion method for the classification technique. The fusion of features creates an overfitting technique and irrelevant features affect the classification performance.

Shobana and Murali [13] applied the skip-gram technique for semantic and contextual feature extraction of words. LSTM model was applied for learning the complex textual pattern and the Adaptive Particle Swarm Optimization (APSO) model was applied for weight parameters. The Adaptive Particle Swarm Optimization model has a local optima trap and lower efficiency in the classification.

2.2. Feature selection methods

Zhang et al. [14] applied the attention model of interactive attributes to consider relevant attributes and interactive relationships to enhance performance for reviews. Multiple interactive attribute encoder deliberates all the attributes to increase the sentiment classification performance. After the local text encoder, multiple interactive attribute encoders were exploited to extract hidden data for text representations of aligning attribute features with self-attention of bilinear interaction. The multi-loss objective function was applied to improve the performance and tested on three datasets such as Amazon, Yelp and IMDB. The developed technique was a supervised model and has a vanishing gradient problem. Gokalp et al. [15] applied Iterated Greedy (IG) technique for feature selection in sentiment classification. The Iterated Greedy model was tested on the Amazon product dataset and displays significant results. The developed technique has limitations of imbalance data problems and overfitting.

2.3. Classification methods

Chen et al. [16] applied sentiment classification techniques to analyze user's review habits to increase the performance of Hierarchical Neural Networks. Based on users, training sets are partitioned in the model and each user review are aggregated that is user's historical reviews. The LSTM-based Hierarchical Neural Network was applied to train target and historical reviews for document representations. The similarities between multiple historical reviews and document representation of target review. The LSTM model has a vanishing gradient problem and the Hierarchical Neural Network model has an overfitting problem in classification.

Zhao et al. [17] applied an attention sharing mechanism and parameter transferring technique for cross-domain sentiment classification that consists of Target Domain Network and Source Domain Network. The pre-training language model was applied as training data in Hierarchical Attention Network such as global vectors for Bidirectional Encoder Representations from Transformers (BERT) and global vectors for word representation. To fine-tune the words, the parameter transferring technique in word and sentence levels was applied from Target Domain Network to Source Domain Network. The system with overfitting degrades the model's performance.

Sazzed and Jayarathna [18] applied a hybrid technique of Self-supervised Sentiment Analyzer that uses lexicon-based technique and machine learning technique for sentiment classification from unlabelled data. The Lexical Rule-based Sentiment Analyzer was applied to predict review semantic orientation for prediction confidence score. The confidence score helps to provide highly accurate pseudo labels for Self-supervised Sentiment Analyzer and this consists of machine learning techniques to improve classification performance for complex reviews and less polarized. The Self-supervised Sentiment Analyzer and Lexical Rule-based Sentiment Analyzer performance was compared with existing unsupervised and achieved higher performance.

Rezapour [19] applied three machine learning classifiers such as Decision Tree, Support Vector Machine (SVM) and Naïve Bayes for sentiment analysis. Sparse word removal was applied for data in the removal of almost sentiments of negative reviews due to positive reviews being applied for reviews majority. Training the model individually for positive and negative reviews helps to increase learning capacity. Maintaining stop words, applications of various N-grams, separate tokenization process, and review title combination with their contents were applied to improve model performance. The learning capacity of the model was less and the feature selection process was not effective.

Kumar et al. [20] applied a hybrid technique of a three-step semi-supervised model that jointly learns aspects and sentiment from review sentences. Each aspect of seed words in a small set was considered and constructed sentiment class for respective semantics in vocabularies of coherent class. The Part-of-Speech (POS) tags were used to construct vocabularies to label sentences subset from training datasets. The semi-automatic technique was applied to label the data and this induced noise in the label during annotation. The overfitting problem in the joint learning method degrades the performance of classification.

Geetha and Renuka [21] performed sentiment analysis on consumer review data to classify positive and negative feelings. Various models such as SVM, LSTM and Naïve Bayes were used for review classification. The deep learning technique of the BERT Base Uncased model was used to solve the problem of sentiment analysis. An improved performance was provided by the BERT model with high accuracy and good prediction than the existing technique. The training of the BERT model generates more weights that create overfitting in classification.

Dadhich and Thankachan [22] perform automatic identification of sentiment analysis using K-Nearest Neighbor, Random Forest, SentiWordNet, Logistic Regression, and Naïve Bayes on English text from products of Flipkart and Amazon. Five key parameters of methods and existing sentiment analysis were presented in this study. The Product Comment Summarizer and Analyzer system was presented in this paper. A generic and automatic comment analyzer was used to find sentiment polarity and provide effective comments. This summarizes and classifies the comments as neutral, negative and positive very effectively. Rakshit et al. [23] applied analytics-based statistical techniques on primary data from samples of millennial samples to effective communication in selling e-marketers on individual human connections. Twitter platform of Twitter data was used for sentiment analysis to analyze the performance of Amazon and Flipkart during Big Billion Day Sales 2019 in India. The Naïve Bayes applied in the model consider the factors are independent and lower efficiency in learning feature importance.

Priyadarshini and Cotton [24] applied LSTM-CNN with grid search of Deep Neural Network. Baseline models including CNN-LSTM, LSTM-CNN, Neural Networks, K-Nearest Neighbors, and CNN were considered in this research. The developed model displays significant improvement, but the overfitting issue disturbs the effectiveness of developed model. Zhang et al. [25] applied a hybrid model for the amazon review dataset and neural networks were applied for the collaborative filtering process. This model applies embedding techniques to perform the model for a smaller number of training data. This model suffers from the limitation of overfitting problems to degrade the performance.

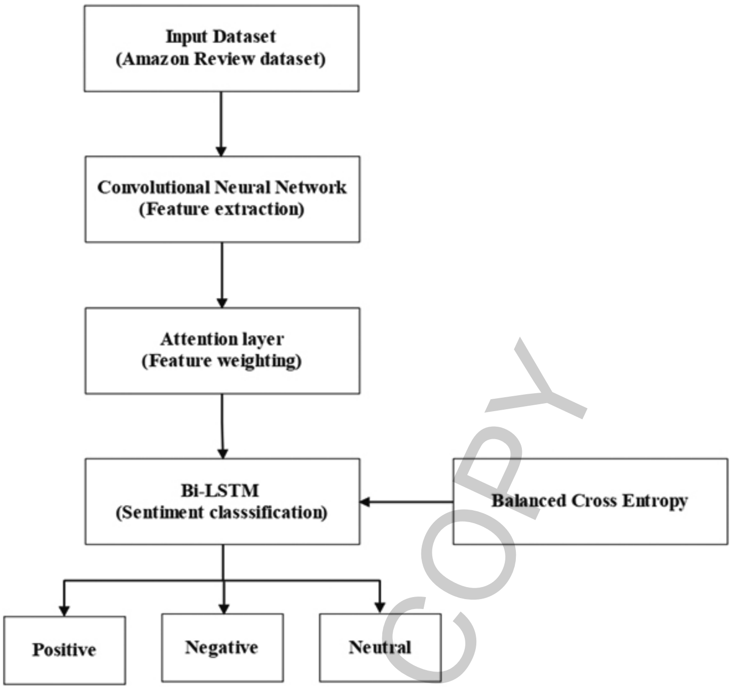

Fig. 1. The flow of Word Embedding Attention-Bi-Directional Long Short-Term Memory (WEA-BiLSTM) model for sentiment classification.

3. Proposed Word Embedding Attention-Bi-directional Long Short-Term Memory (WEA-BiLSTM) model

Here, in this research, the input Amazon Review dataset was used and given to CNN model for feature extraction. The attention layer provides weight value to the features and is applied to the BiLSTM model for sentiment classification. The Balanced Cross-Entropy is applied to maintain gradient in the network. The flow of the WEA-BiLSTM model for sentiment classification is illustrated in Fig. 1.

3.1. Dataset description

This is a sizable crawl of Amazon merchandise reviews. From about 20 million users, this dataset includes 82.83 million unique evaluations. Additionally, it contains 9.35 million pieces, covering the period from May 1996 to July 2014. It contains the meta data: , category, helpfulness votes, item-to-item relationships, price, product image, reviews and ratings, sales Rank, timestamps.

3.2. Word embedding

Word embeddings are calculated in various ways. The placement of each word takes place in a three-dimensional area. The data scientist decides how many variables to use in this area. Also experiment with various dimensions to see which one yields the finest outcome. Dense vectors with a lot less dimension are word embeddings. In addition, the vectors' direction and distance represent the semantic connections between words. The embedding technique tends to give rules to words that are only occasionally encountered in training because it has a larger vocabulary.

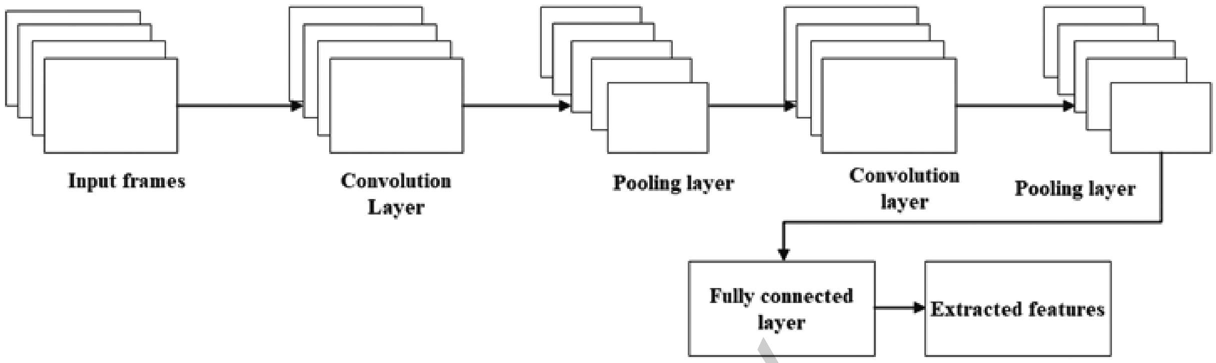

Fig. 2. Architecture of Convolutional Neural Network (CNN) for feature extraction.

Existing pre-trained word embeddings never succeed in sentiment analysis tasks because some sentiment words have similar syntactic and semantic characteristics in the corpus. This study suggests a word embedding technique to enhance the sentence-level emotion classification's efficiency. The mapping between word embeddings and the corresponding sentiment orientations is fully utilized by this technique of sentiment enhancement. Words are first transformed into word embeddings using this technique, and then emotion mapping vectors are applied to each word embedding. Then, word embeddings and the emotion mapping vector that go with them are combined to create sentiment. The predicted sentiment orientations are obtained after reducing the dimensions of sentients through a completely connected layer. Then this output is processed for feature extraction stage where CNN is employed.

3.3. Convolutional Neural Network (CNN)

CNN is excellent at extracting features from images and has been demonstrated to be highly effective at identifying patterns that are challenging to identify using conventional techniques. CNN employs a feature extractor in the training process. The feature extractor used by CNN is made up of unique neural network classes, the weights of which are determined during training. Multiple convolution layers are followed by max-pooling and an activation function in the feature extraction process [26]. A crucial part of the CNN architecture is the convolution layer, which carries out feature extraction. Feature extraction usually entails combining linear and nonlinear operations, such as the convolution operation and the activation function [27]. Humans find the method for comprehending images fascinating, and it is a realistic option for individuals. However, there are more unnoticed difficulties with the machine's ability to comprehend an image. The CNN is a deep learning that aims to mimic the visual technology of an individual. It is stimulated by the visual cortex of the brain. In image processing, which includes detection, localization, segmentation, and classification, among other tasks, CNNs represent a significant advancement.

The major factor driving the model's widespread use is the great CNN effectiveness in classification. Implementing the convolutional layer of learnable weights and biases similar to real neurons is part of the CNN. As demonstrated in Fig. 2, the fundamental components of CNN are convolutional layers, activation functions, and fully linked layers. This article provides a succinct explanation of the CNN, as well as the report includes a thorough description of CNN [28].

Convolutional Layer: The visual system of brain contains neuronal cells that are involved in retrieving image features [29]. Each neural cell extracts different features that aid in the comprehension of an image. Convolutional layers are used to simulate neuronal cells, which allows for the extraction of properties including gradient perception, texture, colors, and edges. Convolutional filters, also known as kernels, have a size of $n \times m \times d$, where d is the image depth, and they are learnable filters in convolutional layers. Following the forward pass, the kernels are convolved throughout the input volume's height and width, as well as the dot product is calculated for filter entries and input. CNN learns which filters to use for texture, color, edge, etc. The output of the convolution layer is used as a layer for the activation function [30].

Activation Function: While non-linearity predominates in actual statistics, non-linear data transformation is applied. This guarantees that specified representations of the input space for diverse output spaces according to the necessities.

This requires real-value number x and converting into 0 and 1. Input with large negative and positive are positioned near 0 and unity, individually which is stated in Eq. (1).

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

In a non-linear function, the real value number x is taken into account and, if x is negative, converts to 0. As demonstrated in Eq. (2), the ReLU activation function is a frequently utilized non-linear function that is faster than those with the sigmoid and $\tanh$ functions and takes less computing time.

$$f(x) = \max(0, x) \tag{2}$$

Pooling: It does a non-linear down-sampling with a combined feature. As a result, dimensional reduction data processing requires less computational resources. To lower the spatial size, data are aggregated based on feature type or space; rotational variance of images outperforms translation and regulates overfitting. A rectangle patch set is created when inputs are separated to use a pooling method. Depending on pooling procedure, a single value is computed to replace every patch. Maximum and average pooling are the two types of pooling that are widely frequently used [31].

Fully Connected Layer: Inputs are interconnected to every node in the subsequent layer and weight values are assigned to every node, similar to a neural network [32]. The model's final result is the sum of the inputs times the respective weights. To carry out the classification task, a fully connected layer is coupled to the sigmoid activation function. In Fig. 3, CNN's completely linked layer is displayed.

3.4. Attention layer

Three region CNN networks are applied in a weighted word representation layer to use knowledge of the text. The word representation is given as input and applied to a convolutional layer built on every tweet. F_n weighted matrix is used to calculate convolutional word vector matrix that is stated as $w \in R^{t \times m}$ regularly to measure local and inherent features and a word vector t is designated in F_n matrix, as in Eq. (3):

$$h_i = f(V_{i:j+t-1} \times W[i] + b_i) \tag{3}$$

where word vector t generates a feature map of $h \in R_{n-t+1}$, bias is b, the matrix's weight is denoted as W, and the non-linearity activation function is denoted as f. Feature maps are produced in convolutional layers, the max-pooling layer is applied to reduce dimension of dataset and extract import features, as displayed in Eq. (4).

$$p_i = \mathbf{\mathit{Max}}[h_i] \tag{4}$$

Here, max-pooling layer is applied in the feature map to denote $P_i \in R^{n-t+\frac{1}{2}}$. Max pooling is utilized to various CNN layers filters to extract correspondingly vital features and not focus on polarity and semantic

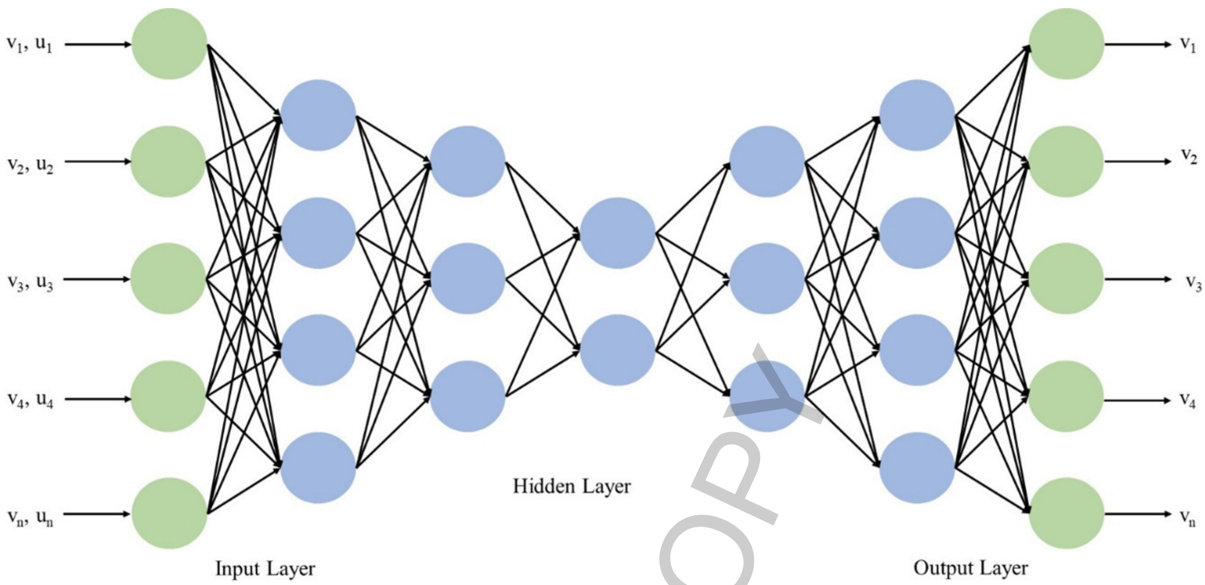

Fig. 3. Convolutional Neural Network (CNN) with fully connected layer.

status. The attention layer is applied to provide the significance of every feature on CNN-generated features, as in Eq. (5).

$$A_i = \frac{\exp(p_i)}{\sum_i \exp(p_i)} \tag{5}$$

The max-pooling attention calculation is given in Eq. (5) above and generated attention score A_i is applied for each feature context p_i.

The attention score output is applied in BiLSTM to learn the feature context. The bi-LSTM model performs sequential maps to generate final features. The final feature context p_i is used for the last feature map from CNN and attention scores are denoted as A_i.

3.5. Bi-directional Long Short-Term Memory (BiLSTM) model

The LSTM model is widely used in sentiment prediction because it provides superior performance on data sequences [33]. The input x_k at the current time, step k and hidden layer output is denoted as h_{k-1} as the previous time step of the hidden layer output. The LSTM cell state c_k and LSTM unit network architecture is shown in Fig. 4. The cell state propagates the gradient and thus LSTM stores the relevant features for the long term than Recurrent Neural Network (RNN) model.

Forget gate f_k is the LSTM model first unit, and information is specified in the input gate to update. Temporarily save the g_k new candidate value before the update a new cell state value. Input is received as x_k and h_{k-1}, the output is calculated with biases, weight parameter and ReLU activation function at each stage. The ReLU activation function overcomes vanishing gradient problem and gates are measured using Eqs (6)–(8):

$$f_k = \sigma(W_x^f x_k + W_h^f h_{k-1} + b^f) \tag{6}$$

$$i_k = \sigma(W_x^i x_k + W_h^i h_{k-1} + b^i) \tag{7}$$

$$g_k = ReLU(W_x^g x_k + W_h^g h_{k-1} + b^g) \tag{8}$$

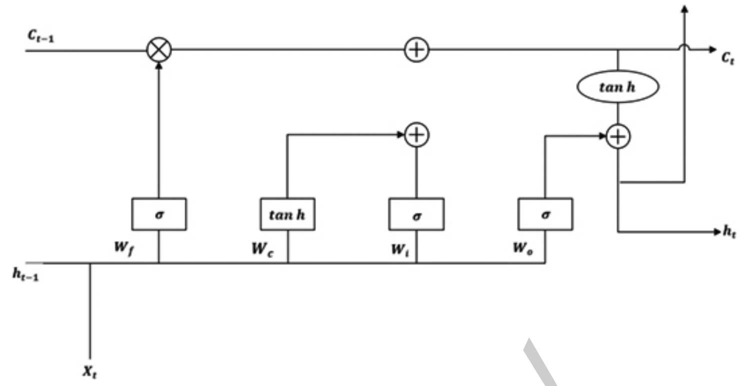

Fig. 4. The Long Short-Term Memory unit consists of input, output and forget gate in Bi-directional Long Short-Term Memory (BiLSTM).

Element-wise product of cell state is applied with forget gate f_k at previous time step c_{k-1}, and new candidate value g_k is applied with the element-wise product for the input gate i_k. Two determined sums is used for cell state c_k, as in Eq. (9).

$$c_k = f_k \odot c_{k-1} + i_k \odot g_k \tag{9}$$

The output gate o_k is the last unit. The output information is determined using this gate. The output gate is calculated similarly to the previous steps, as in Eq. (10).

$$o_k = \sigma(W_x^o x_k + W_h^o h_{k-1} + b^o) \tag{10}$$

Equation (11) is used to compute the hidden state h_t.

$$h_k = o_k \odot \textit{ReLU}(c_k) \tag{11}$$

The prediction is performed using LSTM model, this model uses input as historic capacity data $\{C_k\}$, the size of input and output are the same. The same size of input and output provides input data type and identical output. The LSTM prediction is based on output and input formats.

Two LSTMs layers is applied to learn the sequence of each token-on-token past and future context. The LSTM model manages the order from left to right and another one from vice versa, as in Fig. 5. A hidden forward layer and $\overrightarrow{h}$ hidden unit function at each time step t is calculated in terms of current input step x_t and previous hidden state $\overrightarrow{h}_{t-1}$. Current step input x_t and the future hidden state $\overleftarrow{h}_{t+1}$ are used to compute hidden unit function $\overleftarrow{h}$ in a hidden backward layer. The context representations of forward and backward are used to generate $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$, correspondingly that concatenate with a long vector. The target signals of predictions are based on combined outputs.

The bi-directional network is transformed into high-level sentiment representation in fully connected dense layer to predict text sentiment polarity. Equation (12) provides the output.

$$h_i = \textit{Relu}(w_i h_p + b_i) \tag{12}$$

where BiLSTM network [34,35] provides feature map h_p, the obtained features are h_i, and training learned parameters are w_i, and b_i. The merge feature layer applies sentiment classification performed by the output layer, as shown in Fig. 5. Binary and multi-class datasets are applied with sigmoid and sigmoid classifiers, respectively. Balanced Cross Entropy is utilized to perform difference among actual and predicted text.

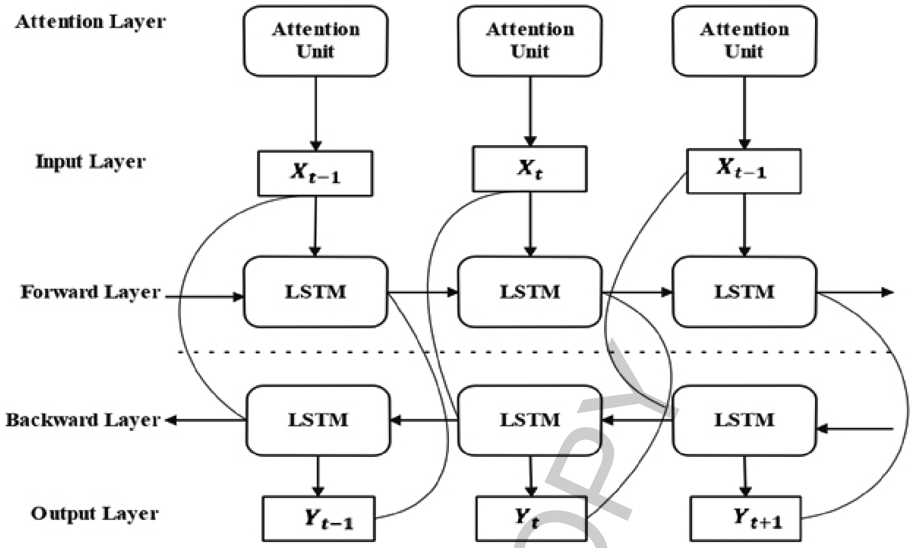

Fig. 5. Attention layer in Bi-directional Long Short-Term Memory (BiLSTM) model for sentiment classification.

3.5.1. Balanced cross-entropy

The deep neural network baseline is combined with two distinct losses of optimization and a small replay memory: the distillation loss and standard softmax cross-entropy.

Model is trained using total loss L, to define weighted distillation loss L_d and standard softmax cross-entropy loss L_c, as in Eq. (13):

$$L = \rho L_d + (1 - \rho) L_c \tag{13}$$

where ρ is applied as N_{t-1}/N_t with incremental step t and N_t as a total number of classes, to balance two losses' importance. At the training stage, the number of new classes are evaluated with a total number of learning classes to update the significance of distillation loss.

At each incremental step t, the new parameters θ_t are used for the first initialization and the previous step parameters θ_{t-1} that is used to maintain previously learned knowledge using distillation loss. The distillation loss L_d for each training sample $(x, y) \in X_t \cup X_M$, as in Eqs (14) and (15):

$$L_d(x) = \sum_{k=1}^{N_{t-1}} -\hat{p}_k(x) \log(p_x(x)) T^2, \tag{14}$$

$$\hat{p}_k(x) = \frac{e^{\frac{\hat{z}_k(x)}{T}}}{\sum_{j=1}^{N_{t-1}} e^{\frac{\hat{z}_j(x)}{T}}}, p_k(x) = \frac{e^{\frac{z_k(x)}{T}}}{\sum_{j=1}^{N_{t-1}} e^{\frac{z_j(x)}{T}}} \tag{15}$$

where associated ground truth label is y, the input data is x, the current model θ_t output logits is $z(x) = [z_1(x), \ldots, z_{N_t}(x)]$, and the previous incremental step θ_{t-1} output logits of model is $z(x) = [\hat{z}_1(x), \ldots, \hat{z}_{N_t} - 1(x)]$.

4. Results

The Word Embedding Attention-Bi-directional Long Short-Term Memory (WEA-BiLSTM) model

Table 1
Performance of WEA-BiLSTM model for various iterations

Iteration	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
0	0	0	0	0
50	91.7	81.4	72.3	76
100	92.5	82.1	73.4	76.5
150	92.8	82.3	73.6	77.2
200	93.4	82.5	73.8	77.8
250	93.7	83.3	74.1	78.1
300	94	84	75.2	78.3
350	94.9	85.7	76	78.7
400	95.6	86	76.1	80.4
450	97	86.8	76.3	80.4
500	97.4	86.8	76.7	81.2

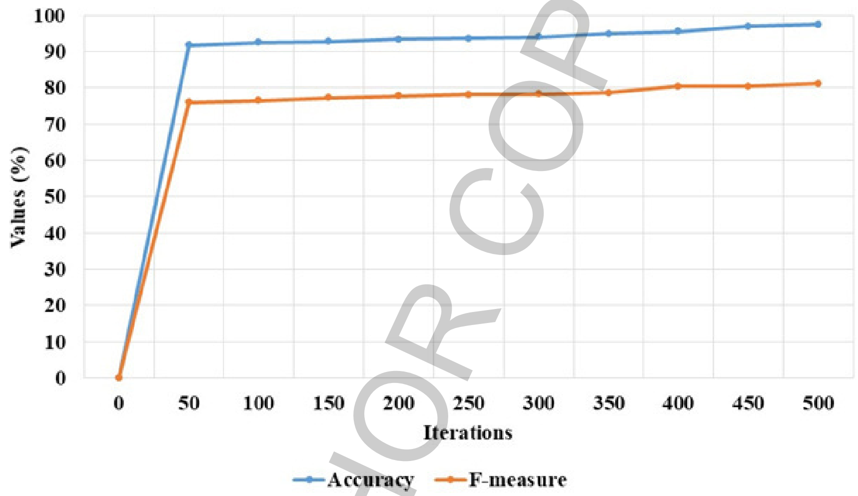

Fig. 6. Accuracy and F-measure of WEA-BiLSTM model for various iterations.

is applied on the Amazon dataset for the sentiment analysis and compared with other methods. The performance and error metrics were measured on the results of WEA-BiLSTM model for comparison.

The WEA-BiLSTM model is applied for sentiment analysis and its performance is measured for various iterations, as shown in Table 1. This shows that the model has higher accuracy due to its capacity to detect sentiment. The attention technique uses word embedding for the convolution process which helps to provide a better feature representation. The recall value of the model is low due to the presence of text information related to sarcasm and distinguishing features. In 400th iteration, the WEA-BiLSTM model accuracy is 95.6%, precision is 86%, recall is 76%, and F-Measure is 78.7%.

The WEA-BiLSTM model accuracy and F-measure are measured in sentiment analysis for various iterations, as shown in Fig. 6. This shows that the WEA-BiLSTM model has higher accuracy and considerably lower performance than F-Measure. The WEA-BiLSTM model has lower performance in the negative class and this reduces the F-measure of the model. The model has the lesser results in the negative class since the presence of sarcastic information from the user.

The precision and recall of the WEA-BiLSTM model is measured for various iterations, as shown in Fig. 7. The WEA-BiLSTM model achieves higher precision and recall values within 50 iterations. The WEA technique increases the learning rate of the BiLSTM model based on weight values provided to

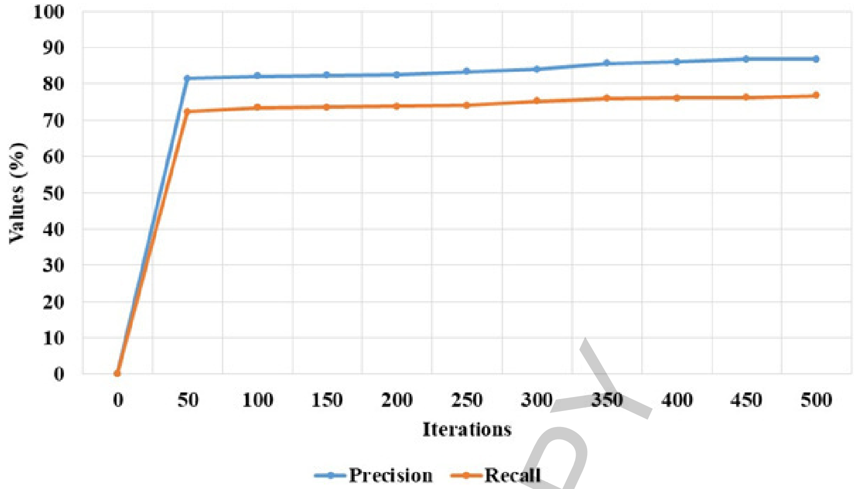

Fig. 7. Precision and recall of WEA-BiLSTM model for various iterations.

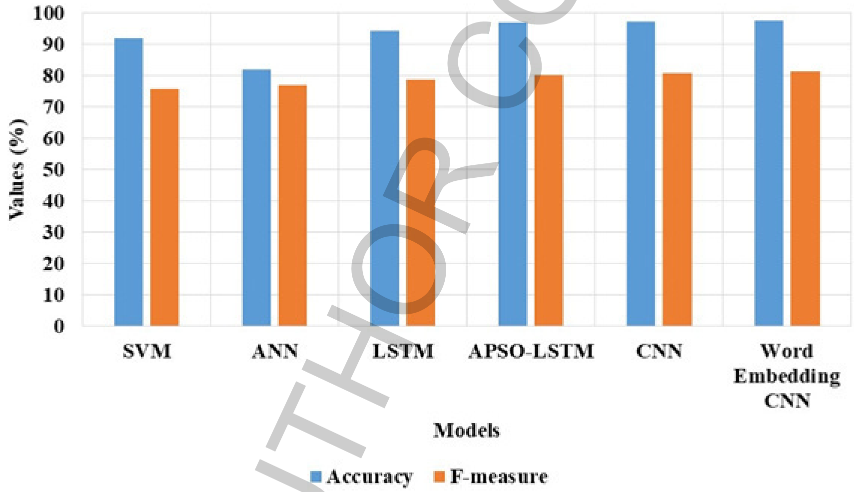

Fig. 8. The WEA-BiLSTM model classifier comparison on sentiment analysis.

input text. The WEA technique provides higher weight to the words having strong relation with class and this helps to enhance the model effectiveness.

The performance of the WEA-BiLSTM model is evaluated with other classifiers, as displayed in Table 2 and Fig. 8. WEA with CNN can generate deep features and provide higher weight to the words having strong relation with class. The existing SVM model has an imbalance data problem, and the LSTM model has a vanishing gradient problem. The APSO-LSTM model has a local optima trap and misses some potential features for the classification. The CNN model has an overfitting issue since the formation of additional features in extraction process. The WEA-BiLSTM model has 97.4% accuracy, and 86.8% precision, and existing CNN contains 97.1% accuracy, and 85.4% precision in sentiment analysis.

The precision and recall value of the WEA-BiLSTM model is calculated and compared with existing classifiers on sentiment analysis, as shown in Fig. 9. The WEA-BiLSTM model uses word embedding to provide a higher weight value to the words having strong relation with dataset categories. This increases

Table 2
The WEA-BiLSTM classifier comparison on sentiment analysis

ANN: Artificial Neural Network
APSO: Adaptive Particle Swarm Optimization
CNN: Convolutional Neural Network
LSTM: Long Short-Term Memory
SVM: Support Vector Machine

Comparison	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
SVM	91.8	81	73	75.6
ANN	81.9	71.9	72	76.8
LSTM	94.1	83	74.5	78.57
APSO-LSTM	96.8	85.28	76.08	80.04
CNN	97.1	85.6	76.33	80.64
Word Embedding CNN	97.4	86.8	76.7	81.2

Table 3
The WEA-BiLSTM model performance for various categories

Category	Methods	Accuracy (%)	AUC	Precision (%)	Recall (%)	F-measure (%)
Books	Iterated Greedy (IG)	88	0.93	88	88	88
	CNN	90	0.94	90	90	90
	Proposed CNN	93	0.96	92	92	92
DVD	Iterated Greedy (IG)	90.25	0.94	90	90	90
	CNN	91.4	0.95	90	90	90
	Proposed CNN	93.5	0.97	93	93	93
Electronics	Iterated Greedy (IG)	91.63	0.95	92	92	92
	CNN	92.3	0.96	93	93	93
	Proposed CNN	93.7	0.97	95	95	95
Kitchen	Iterated Greedy (IG)	93.09	0.96	93	93	93
	CNN	94.1	0.97	93	93	93
	Proposed CNN	95.2	0.98	94	94	94
AVG	Iterated Greedy (IG)	90.74	0.94	90.75	90.75	90.75
	CNN	91.95	0.95	91.5	91.5	91.5
	Proposed CNN	93.85	0.97	93.5	93.5	93.5

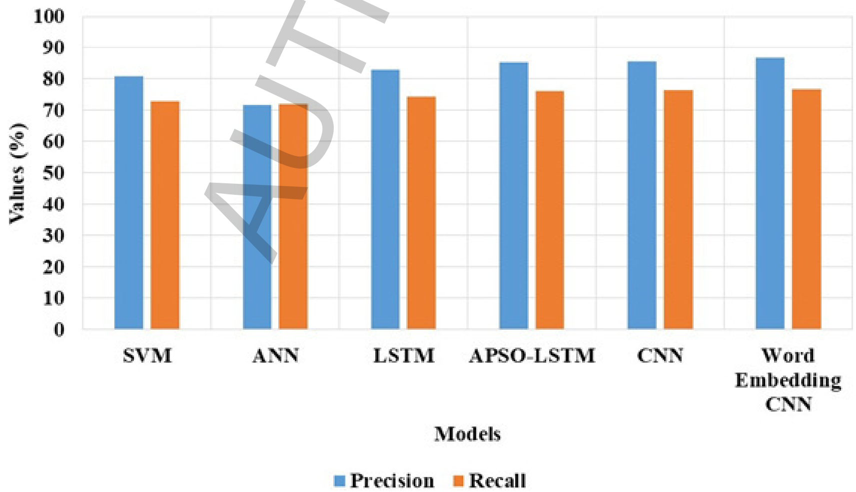

Fig. 9. WEA-BiLSTM precision and recall on sentiment analysis.

Table 4
WEA-BiLSTM model Mean Absolute Error (MAE) on sentiment analysis

A^3NCF: aspect attention-based neural collaborative filtering
ItemKNN: item-based k-nearest neighbors' algorithm
SlopeOne: simplest form of non-trivial item-based collaborative filtering
SVD: singular value decomposition for matrix factorization
UTER: unified text/paragraph embeddings of user reviews and item descriptions [25]

Datasets	SlopeOne	ItemKNN	SVD	A3NCF	UTER	CNN	Proposed CNN
Baby	0.8557	0.8536	0.8123	0.8075	0.7966	0.7854	0.7704
Sports and Outdoor	0.663	0.6771	0.6374	0.633	0.6169	0.6087	0.5843
Pet Supplies	0.8917	0.9202	0.88	0.8746	0.855	0.8542	0.8412
Video Games	0.8251	0.8482	0.7935	0.7916	0.7907	0.7891	0.77

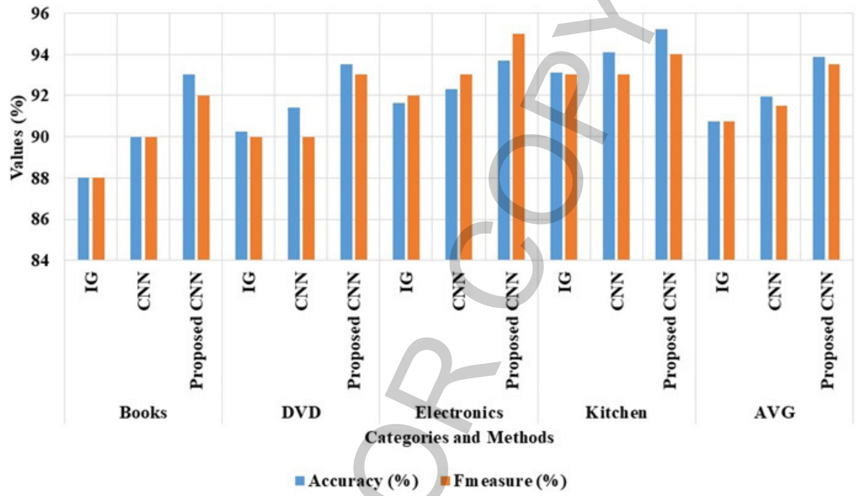

Fig. 10. The WEA-BiLSTM model accuracy and F-measure for several categories.

WEA-BiLSTM model performance related to the product category and increases the precision and recall value. The existing CNN model has lower efficiency since the formation of additional features in the convolutional layer. LSTM has a vanishing gradient problem and this degrades the model performance.

The WEA-BiLSTM model performance is evaluated on the parameters of accuracy and F-measure for several categories in the dataset, as exposed in Table 3 and Fig. 10. The WEA-BiLSTM contains higher performance in every category in the dataset. The WEA-BiLSTM model applies the word embedding technique to provide higher weight value to words having strong relation with classes. The CNN extracted features and weight value increases the learning rate of the LSTM model in classification. The balanced cross-entropy maintains the gradient values in the model that helps to overcome the vanishing gradient problem. The WEA-BiLSTM model has 93.85% accuracy, 93.5% precision, and CNN model has 91.95% accuracy, and 91.5% precision.

The precision and recall value of WEA-BiLSTM model for various categories in the dataset is shown in Fig. 11. The WEA-BiLSTM model has higher precision and recall value than existing techniques in sentiment analysis. The WEA-BiLSTM model applies a higher weight value for the words having strong relation with classes. This helps to increase the performance of the model class-wise, thus increasing the precision and recall value.

The Mean Absolute Error (MAE) value of WEA-BiLSTM model for various categories on the Amazon dataset is calculated and evaluated by conventional methods, as displayed in Table 4 and Fig. 12. WEA-

Table 5
WEA-BiLSTM Root Mean Square Error (RMSE) on sentiment analysis

Datasets	SlopeOne	ItemKNN	SVD	A3NCF	UTER	CNN	Proposed CNN
Baby	1.2012	1.1854	1.0704	1.0324	1.0146	0.9987	0.9862
Sports and Outdoor	1	0.9959	0.9065	0.8548	0.8501	0.8421	0.8369
Pet Supplies	1.2454	1.2658	1.1448	1.1058	1.0627	0.9965	0.9921
Video Games	1.1421	1.1634	1.0511	1.0269	1.0164	0.9988	0.9908

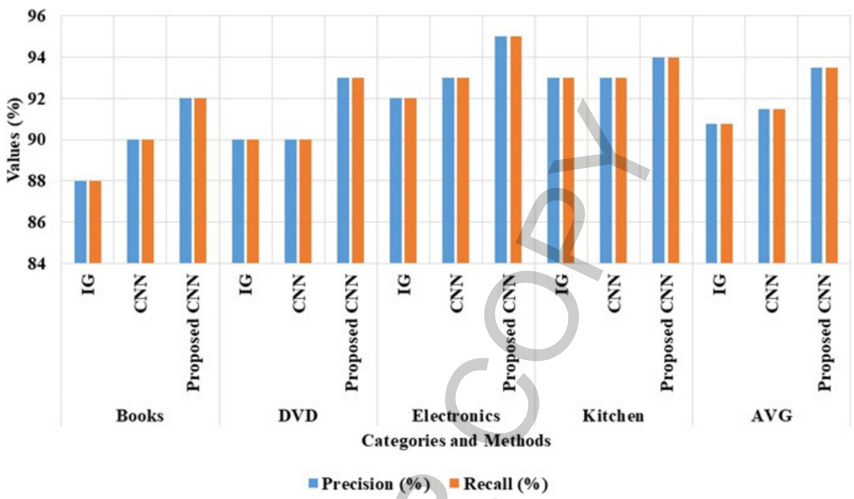

Fig. 11. The WEA-BiLSTM model precision and recall for various categories.

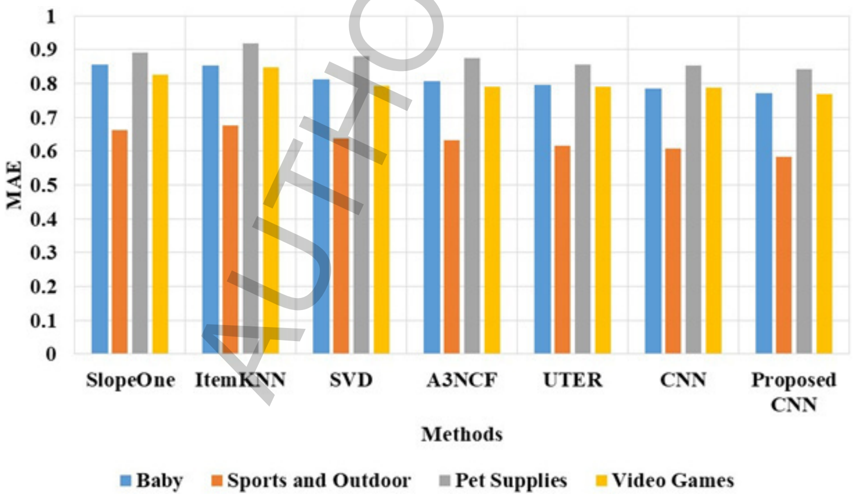

Fig. 12. Mean Absolute Error (MAE) of WEA-BiLSTM model on sentiment analysis.

BiLSTM contains higher performance than existing method. The WEA technique provides a higher weight value to the words having strong relation with classes. The balanced cross-entropy maintains the gradient value in the network and avoids the vanishing gradient problem.

The WEA-BiLSTM model measures Root Mean Square Error (RMSE) value for various categories on the dataset, as displayed in Table 5 and Fig. 13. The WEA-BiLSTM shows better class-wise output due to

Table 6
WEA-BiLSTM model for sentiment analysis

Algorithm	User cold start		Item cold start		User-item cold start	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
CNN	0.8108	1.1984	0.8065	1.1982	0.8057	1.2008
UTER	0.8123	1.2036	0.8106	1.2033	0.8112	1.2089
A3NCF	0.8274	1.2137	0.8313	1.221	0.8336	1.2227
SVD	0.8409	1.2235	0.8446	1.2252	0.848	1.2276
ItemKNN	0.8653	1.251	0.8714	1.2501	0.8752	1.2532
SlopeOne	0.8726	1.2626	0.8805	1.2639	0.8876	1.2683
Word Embedding CNN	0.8044	1.1920	0.7996	1.1905	0.8003	1.1941

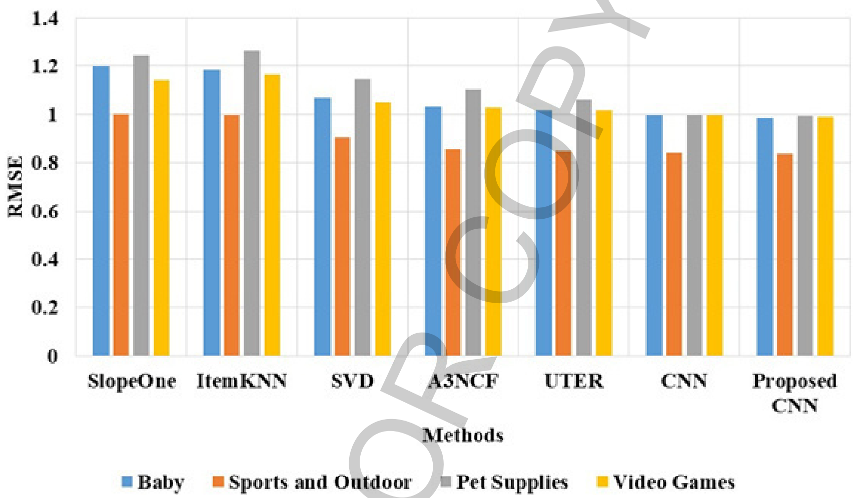

Fig. 13. Root Mean Square Error (RMSE) of WEA-BiLSTM model on sentiment analysis.

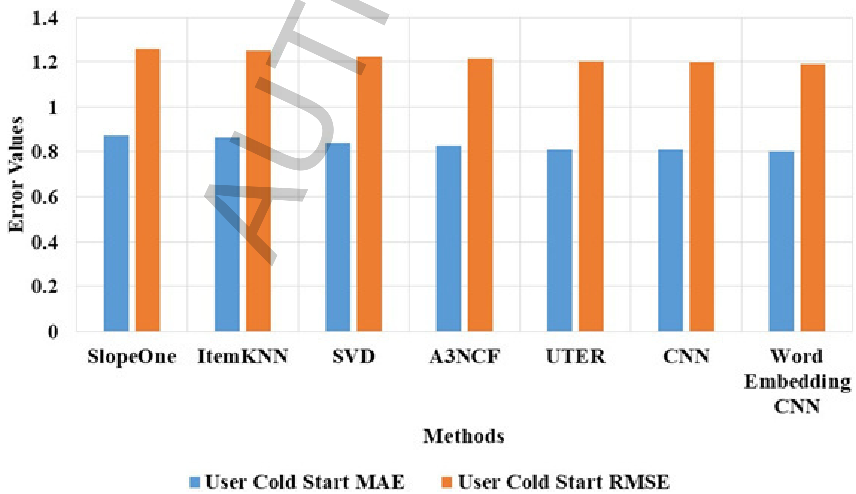

Fig. 14. The WEA-BiLSTM model for user cold start.

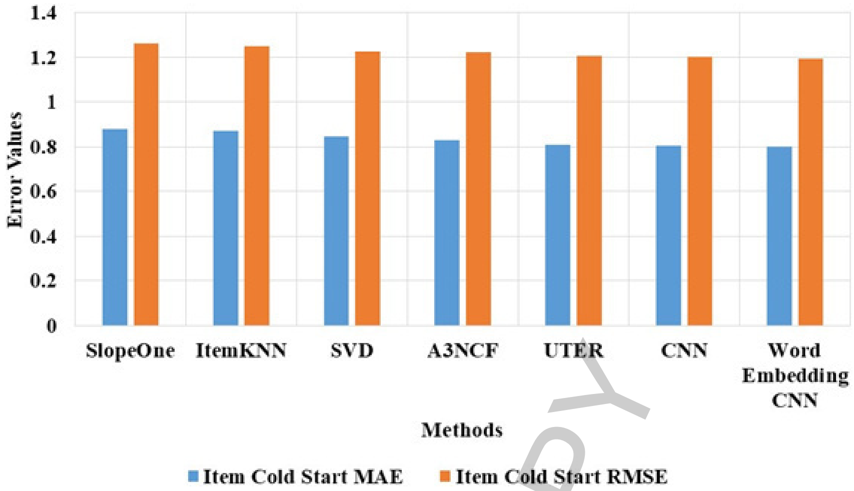

Fig. 15. The WEA-BiLSTM model for item cold start.

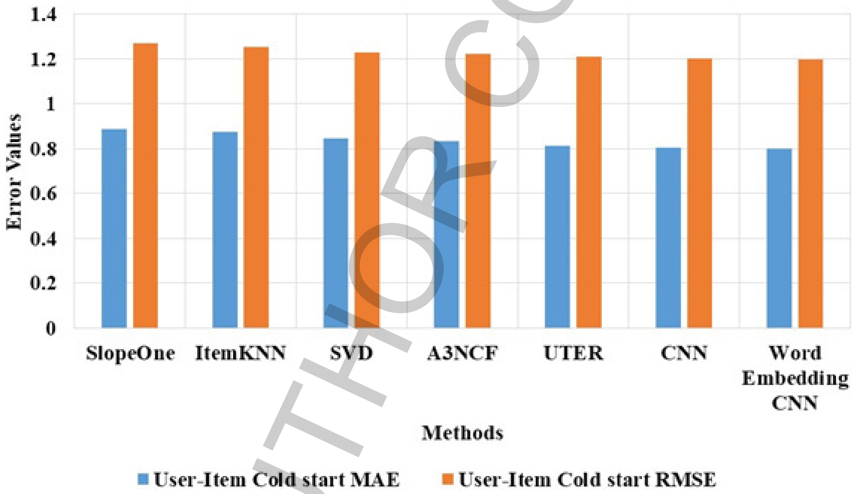

Fig. 16. WEA-BiLSTM model for user Item cold start.

weight value which is provided related to product category. The conventional CNN consist of overfitting problems due to generation of more features in convolutional layers.

The WEA-BiLSTM model is evaluated with Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values for various cold start problems, as shown in Table 6. The user cold start has less information about the user information and the item cold start has less information about products. The WEA-BiLSTM model uses the CNN based feature extraction that helps to extract deep features and provide higher weight values to the words having strong relation.

The WEA-BiLSTM model performance is measured with Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for various user, as shown in Figs 14–16, respectively. The WEA-BiLSTM achieved supreme results in classification over conventional models on sentiment analysis. The WEA-BiLSTM model has higher efficiency for a smaller number of training data due to the model uses CNN feature extraction for classification.

5. Conclusion

Sentiment analysis on review provides the sentiment of the user about the product that is useful for product development. The existing LSTM and CNN based models were applied in sentiment analysis that has limitations of vanishing gradient problem and overfitting problem in classification. This study proposed the WEA technique to provide higher weight values to the words having strong relation with classes. The CNN model is applied for feature extraction due to the generation of more features in the convolutional layer. The Balanced Cross Entropy is applied to maintain the gradient in the network to solve the vanishing gradient problem. The CNN feature extraction helps to provide higher performance for a smaller number of training data for classification. The future work comprises applying the feature selection method to select relevant features and achieve higher performance in imbalance data.

Funding

This research received no external funding.

Conflict of interest

The authors declare that they have no conflict of interest.

Data availability

The datasets generated during and/or analysed during the current study are available in the [Amazon Review datasets] repository: [https://www.kaggle.com/datasets/bittlingmayer/amazonreviews].

References

[1] N.M. Alharbi, N.S. Alghamdi, E.H. Alkhammash and J.F. Al Amri, Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews, Mathematical Problems in Engineering, 2021, 5536560. doi: 10.1155/2021/5536560.

[2] M. Sivakumar and S.R. Uyyala, Aspect-based sentiment analysis of mobile phone reviews using LSTM and fuzzy logic, *International Journal of Data Science and Analytics* **12**(4) (2021), 355–367. doi: 10.1007/s41060-021-00277-x.

[3] J. Liu, S. Zheng, G. Xu and M. Lin, Cross-domain sentiment aware word embeddings for review sentiment analysis, *International Journal of Machine Learning and Cybernetics* **12**(2) (2021), 343–354. doi: 10.1007/s13042-020-01175-7.

[4] I. Chatterjee, M. Zhou, A. Abusorrah, K. Sedraoui and A. Alabdulwahab, Statistics-based outlier detection and correction method for amazon customer reviews, *Entropy* **23**(2) (2021), 1645. doi: 10.3390/e23121645.

[5] A. Alsayat, Improving sentiment analysis for social media applications using an ensemble deep learning language model, *Arabian Journal for Science and Engineering* **47**(2) (2022), 2499–2511. doi: 10.1007/s13369-021-06227-w.

[6] M. Birjali, M. Kasri and A. Beni-Hssane, A comprehensive survey on sentiment analysis: Approaches, challenges and trends, *Knowledge-Based Systems* **226** (2021), 107134. doi: 10.1016/j.knosys.2021.107134.

[7] M.E. Basiri, S. Nemati, M. Abdar, E. Cambria and U.R. Acharya, ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis, *Future Generation Computer Systems* **115** (2021), 279–294. doi: 10.1016/j.future.2020.08.005.

[8] C.N. Dang, M.N. Moreno-García and F. De la Prieta, Using Hybrid Deep Learning Models of Sentiment Analysis and Item Genres in Recommender Systems for Streaming Services, *Electronics* **10**(20) (2021), 2459. doi: 10.3390/electronics10202459.

[9] M.M. Agüero-Torales, J.I. Abreu Salas and A.G. López-Herrera, Deep learning and multilingual sentiment analysis on social media data: An overview, *Applied Soft Computing* **107** (2021), 107373. doi: 10.1016/j.asoc.2021.107373.

[10] M. Singh, A.K. Jakhar and S. Pandey, Sentiment analysis on the impact of coronavirus in social life using the BERT model, *Social Network Analysis and Mining* **11**(1) (2021), 33. doi: 10.1007/s13278-021-00737-z.

[11] S.K. Banbhrani, B. Xu, H. Lin and D.K. Sajnani, Spider Taylor-ChOA: Optimized deep learning based sentiment classification for review rating prediction, *Applied Sciences* **12**(7) (2022), 3211. doi: 10.3390/app12073211.

[12] S. Wei and S. Song, Sentiment classification of tourism reviews based on visual and textual multifeature fusion, *Wireless Communications and Mobile Computing* **2022** (2022), 1–10. doi: 10.1155/2022/9940817.

[13] J. Shobana and M. Murali, An efficient sentiment analysis methodology based on long short-term memory networks, *Complex and Intelligent Systems* **7**(5) (2021), 2485–2501.

[14] Y. Zhang, J. Wang and X. Zhang, Personalized sentiment classification of customer reviews via an interactive attributes attention model, *Knowledge-Based Systems* **226** (2021), 107135. doi: 10.1016/j.knosys.2021.107135.

[15] O. Gokalp, E. Tasci and A. Ugur, A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification, *Expert Systems with Applications* **146** (2020), 113176. doi: 10.1016/j.eswa.2020.113176.

[16] J. Chen, J. Yu, S. Zhao and Y. Zhang, User's review habits enhanced hierarchical neural network for document-level sentiment classification, *Neural Processing Letters* **53**(3) (2021), 2095–2111. doi: 10.1007/s11063-021-10423-y.

[17] C. Zhao, S. Wang, D. Li, X. Liu, X. Yang and J. Liu, Cross-domain sentiment classification via parameter transferring and attention sharing mechanism, *Information Sciences* **578** (2021), 281–296. doi: 10.1016/j.ins.2021.07.001.

[18] S. Sazzed and S. Jayarathna, SSentiA: A Self-supervised Sentiment Analyzer for classification from unlabeled data, *Machine Learning with Applications* **4** (2021), 100026. doi: 10.1016/j.mlwa.2021.100026.

[19] M. Rezapour, Sentiment classification of skewed shoppers' reviews using machine learning techniques, examining the textual features, *Engineering Reports* **3**(1) (2021), e12280. doi: 10.1002/eng2.12280.

[20] A. Kumar, P. Gupta, R. Balan, L.B.M. Neti and A. Malapati, BERT based semi-supervised hybrid approach for aspect and sentiment classification, *Neural Processing Letters* **53**(6) (2021), 4207–4224. doi: 10.1007/s11063-021-10596-6.

[21] M.P. Geetha and D. Karthika Renuka, Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model, *International Journal of Intelligent Networks* **2** (2021), 64–69. doi: 10.1016/j.ijin.2021.06.005.

[22] A. Dadhich and B. Thankachan, Social and Juristic challenges of AI for Opinion Mining Approaches on Amazon and Flipkart Product Reviews Using Machine Learning Algorithms, *SN Computer Science* **2**(3) (2021), 180. doi: 10.1007/s42979-021-00554-3.

[23] P. Rakshit, P.K. Srivastava, M. Afjal and S.K. Srivastava, Sentimental Analytics on Indian Big Billion Day of Flip Kart and Amazon, *SN Computer Science* **2**(3) (2021), 204. doi: 10.1007/s42979-020-00441-3.

[24] I. Priyadarshini and C. Cotton, A novel LSTM-CNN-grid search-based deep neural network for sentiment analysis, *The Journal of Supercomputing* **77**(12) (2021), 13911–13932. doi: 10.1007/s11227-021-03838-w.

[25] Y. Zhang, Z. Liu and C. Sang, Unifying paragraph embeddings and neural collaborative filtering for hybrid recommendation, *Applied Soft Computing* **106** (2021), 107345. doi: 10.1016/j.asoc.2021.107345.

[26] J. Lu, L. Tan and H. Jiang, Review on convolutional neural network (CNN) applied to plant leaf disease classification, *Agriculture* **11**(8) (2021), 707. doi: 10.3390/agriculture11080707.

[27] Y.-D. Zhang, S.C. Satapathy, D.S. Guttery, J.M. Górriz and S.-H. Wang, Improved breast cancer classification through combining graph convolutional network and convolutional neural network, *Information Processing and Management* **58**(2) (2021), 102439. doi: 10.1016/j.ipm.2020.102439.

[28] M. Niu, Y. Lin and Q. Zou, sgRNACNN: Identifying sgRNA on-target activity in four crops using ensembles of convolutional neural networks, *Plant Molecular Biology* **105**(4) (2021), 483–495. doi: 10.1007/s11103-020-01102-y.

[29] H. Song, x.-y. Han, C.E. Montenegro-Marin and S. krishnamoorthy, Secure prediction and assessment of sports injuries using deep learning based convolutional neural network, *Journal of Ambient Intelligence and Humanized Computing* **12**(3) (2021), 3399–3410. doi: 10.1007/s12652-020-02560-4.

[30] W. Chen, M. Jiang, W.-G. Zhang and Z. Chen, A novel graph convolutional feature based convolutional neural network for stock trend prediction, *Information Sciences* **556** (2021), 67–94. doi: 10.1016/j.ins.2020.12.068.

[31] H.D. Nguyen, K.P. Tran, S. Thomassey and M. Hamad, Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management, *International Journal of Information Management* **57** (2021), 102282. doi: 10.1016/j.ijinfomgt.2020.102282.

[32] J. Cao, Z. Li and J. Li, Financial time series forecasting model based on CEEMDAN and LSTM, *Physica A: Statistical Mechanics and Its Applications* **519** (2019), 127–139. doi: 10.1016/j.physa.2018.11.061.

[33] K. Park, Y. Choi, W.J. Choi, H.-Y. Ryu and H. Kim, LSTM-based battery remaining useful life prediction with multi-channel charging profiles, *IEEE Access* **8** (2020), 20786–20798. doi: 10.1109/ACCESS.2020.2968939.

[34] M.F. Aslan, M.F. Unlersen, K. Sabanci and A. Durdu, CNN-based transfer learning-BiLSTM network: A novel approach for COVID-19 infection detection, *Applied Soft Computing* **98** (2021), 106912. doi: 10.1016/j.asoc.2020.106912.

[35] S.K. Challa, A. Kumar and V.B. Semwal, A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data, The Visual Computer, 2021. doi: 10.1007/s00371-021-02283-3.

Author's Bios

Vijaya Sagvekar is currently pursuing PhD (Computer Engineering) at PAHER University, Udaipur (Rajasthan) and also working as an Assistant Professor in Information technology department at Vasantdada Patil college of engineering and visual arts, Mumbai. She has teaching, research experience of 16 years. Her area of interest are Database management systems, Software engineering, Web Mining, Data Optimization, Network security.

Prashant Sharma working as an Associate Professor & HOD in Department of Computer Science and Engineering at Pacific (PAHER) University, Udaipur (Rajasthan), India. He has above 9 years teaching experience. He completed PhD (CSE) in 2018 and his research interest area is Artificial Intelligence, Machine Learning and Nature Inspired Algorithms.

पेटेंट कार्यालय
शासकीय जर्नल

OFFICIAL JOURNAL
OF
THE PATENT OFFICE

निर्गमन सं. 06/2023	शुक्रवार	दिनांक: 10/02/2023
ISSUE NO. 06/2023	**FRIDAY**	**DATE: 10/02/2023**

पेटेंट कार्यालय का एक प्रकाशन
PUBLICATION OF THE PATENT OFFICE

(12) PATENT APPLICATION PUBLICATION

(19) INDIA

(22) Date of filing of Application :26/01/2023

(21) Application No.202311005260 A

(43) Publication Date : 10/02/2023

(54) Title of the invention : PREDICTING OVERALL CUSTOMER SATISFACTION FOR AN EFFECTIVE PRODUCT OPINION ANALYSIS ON E-COMMERCE WEBSITES

(51) International classification	:G06Q0030060000, G06Q0030020000, G06Q0050000000, G06Q0030000000, G06F0040300000
(86) International Application No Filing Date	:NA :NA
(87) International Publication No	: NA
(61) Patent of Addition to Application Number Filing Date	:NA :NA
(62) Divisional to Application Number Filing Date	:NA :NA

(71)**Name of Applicant :**

1)Ms Vijaya Sagvekar

Address of Applicant :Research Scholar, Computer Engineering, Pacific Academy Of Higher Education And Research University Pacific Hills, PratapnagarExtn., Airport Road, Debari, Udaipur – 313003 Debari ----------- -----------

2)Dr. Prashant Sharma

3)Dr. Pradip Mane

4)Mrs. Kalpana Prasanna Deorukkhar

5)Vasantdada Patil Pratishthan's College of Engineering & Visual Arts

Name of Applicant : NA

Address of Applicant : NA

(72)**Name of Inventor :**

1)Ms Vijaya Sagvekar

Address of Applicant :Research Scholar, Computer Engineering, Pacific Academy Of Higher Education And Research University Pacific Hills, PratapnagarExtn., Airport Road, Debari, Udaipur – 313003 Debari ----------- -----------

2)Dr. Prashant Sharma

Address of Applicant :Associate Professor Department of Computer Science and Engineering Geetanjali Institute of Technical Studies, Udaipur, Rajasthan Dabok ----------- -----------

3)Dr. Pradip Mane

Address of Applicant :Associate Professor & HOD Information technology department, Vasantdada Patil Pratishthan's College of Engineering & Visual Arts, Vasantdada Patil Educational Complex, Eastern Express Highway, Padmabhushan Vasantdada Patil Marg, Sion Mumbai 400022. Sion ----------- -----------

4)Mrs. Kalpana Prasanna Deorukkhar

Address of Applicant :Assistant Professor, Computer Engineering department, Fr. Conceicao Rodrigues College of Engineering, Fr. Agnel Ashram, Bandstand Promenade, Mount Mary, Bandra West, Mumbai, Maharashtra Bandra ----------- -----------

(57) Abstract :

The present invention is Predicting Overall Customer Satisfaction for an Effective Product Opinion Analysis on E-Commerce Websites E-commerce websites have been emerged in a high range of marketing benefits for the users to publish or share the experience of the received product by posting review that contain useful comments, opinions and feedback on the product. Nowadays, millions of customers gain opportunities to compare similar products in online websites and pick their favourites in digital retailers, such as Amazon.com and Taobao.com. Customer reviews in social media and electronic commerce Websites contain valuable electronic word information of products. In the early process, customers were not having any information about the previous customer reviews but made the decisions for their purchase on the expectations about the product quality. Secondly, the customers observe an average customers value who was the early customers for that product helps to make better decision. Similarly, the sentiment analysis in one of the current researches where the opinions of the people on the product analyzed the written reviews in text format.

No. of Pages : 10 No. of Claims : 4

CERTIFICATES

DEPARTMENT OF PG STUDIES AND DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(Pacific Academy of Higher Education and Research University, Udaipur)

INTERNATIONAL VIRTUAL CONFERENCE
on
EMERGING ERA OF APPLICATIONS OF COMPUTER

Certificate Awarded to

This is to certify that *Vijaya Ravindra Sagvekar*

of *Paher Udaipur Rajsthan*

has Presented Paper entitled *Product opinion analysis for customer satisfaction on E-Commerce Websites*

in this International Virtual Conference on 15th-16th January 2022 held at Pacific Academy of Higher Education and Research University, Udaipur.

Prof. Hemant Kothari	Dr. Dilendra Hiran	Dr. Tanveer Ahmed Kazi	Dr. Prashant Sharma
Conference Chairman	Conference Co-Chairman	Conference Director	Conference Organizing Secretary
PAHER University, Udaipur	PAHER University, Udaipur	PAHER University, Udaipur	PAHER University, Udaipur

Sanjivani Rural Education Society's

Sanjivani College of Engineering, Kopargaon

(An Autonomous Institute, Affiliated to SPPU, Approved by AICTE)

(Accredited by NAAC with 'A' Grade)

DEPARTMENT OF COMPUTER ENGINEERING

(Accredited by NBA)

IOS Press Springer

SCOE/COMP/ICRTC2021/P63

International Conference

on

RECENT TRENDS IN COMPUTING (ICRTC 2021)

Certificate

This is to certify that Ms. VIJAYA SAGVEKAR, Pacific University Udaipur, has participated / presented a paper in the "International Conference on Recent Trends in Computing (ICRTC 2021)" organized by Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India during May 21 & 22, 2021.

Title of the paper: STUDY ON PRODUCT OPINION ANALYSIS FOR CUSTOMER SATISFACTION ON E-COMMERCE WEBSITES

Dr. D. B. Kshirsagar
[Convenor & HoD-Computer]

Verify @

Dr. A.G. Thakur
Director

PLAGIARISM REPORT

Predicting overall customer satisfaction for an effective opinion analysis on E-commerce websites

ORIGINALITY REPORT

4%	2%	3%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	1%
2	link.springer.com Internet Source	1%
3	Yihao Zhang, Zhi Liu, Chunyan Sang. "Unifying paragraph embeddings and neural collaborative filtering for hybrid recommendation", Applied Soft Computing, 2021 Publication	<1%
4	Behrooz Noori. "Classification of Customer Reviews Using Machine Learning Algorithms", Applied Artificial Intelligence, 2021 Publication	<1%
5	Marjan Kamyab, Guohua Liu, Michael Adjeisah. "Attention-Based CNN and Bi-LSTM Model Based on TF-IDF and GloVe Word Embedding for Sentiment Analysis", Applied Sciences, 2021 Publication	<1%

6	www.mdpi.com Internet Source	<1 %
7	www.coursehero.com Internet Source	<1 %
8	programminglanguages084.blogspot.com Internet Source	<1 %
9	Soumya S., Pramod K.V.. "Sentiment analysis of malayalam tweets using machine learning techniques", ICT Express, 2020 Publication	<1 %
10	"Data Science", Springer Science and Business Media LLC, 2018 Publication	<1 %
11	Submitted to Deakin University Student Paper	<1 %
12	docplayer.net Internet Source	<1 %
13	dspace.thapar.edu:8080 Internet Source	<1 %
14	www.semanticscholar.org Internet Source	<1 %
15	research.wur.nl Internet Source	<1 %
16	"Proceedings of the International Conference on Advanced Intelligent Systems and	<1 %

Informatics 2020", Springer Science and Business Media LLC, 2021
Publication

17	Md Rafiqul Islam, Shaowu Liu, Xianzhi Wang, Guandong Xu. "Deep learning for misinformation detection on online social networks: a survey and new perspectives", Social Network Analysis and Mining, 2020 Publication	<1%
18	Pramod Sunagar, Darshana A Naik, V Sangeetha, Anita Kanavalli, S Seema. "Feedback Collection and Sentiment Analysis on the Product Reviews", 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), 2021 Publication	<1%
19	J. Shobana, M. Murali. "An efficient sentiment analysis methodology based on long short-term memory networks", Complex & Intelligent Systems, 2021 Publication	<1%
20	Submitted to University of Hertfordshire Student Paper	<1%
21	ebin.pub Internet Source	<1%
22	deepblue.lib.umich.edu Internet Source	

<1%

23 "Machine Learning for Cyber Security", Springer Science and Business Media LLC, 2020 <1%
Publication

24 Mayur Wankhade, Annavarapu Chandra Sekhara Rao, Chaitanya Kulkarni. "A survey on sentiment analysis methods, applications, and challenges", Artificial Intelligence Review, 2022 <1%
Publication

25 arxiv.org <1%
Internet Source

26 "Natural Language Processing and Chinese Computing", Springer Science and Business Media LLC, 2018 <1%
Publication

27 M. Jagadeesan, T.M Saravanan, P.A. Selvaraj, S. Nithya, K. Snegha, P. Sowmiya. "Social Media News Classification Using Machine Learning Algorithms", 2023 International Conference on Computer Communication and Informatics (ICCCI), 2023 <1%
Publication

28 wireilla.com <1%
Internet Source

Recommender Systems Handbook, 2015.

29	Publication		<1 %
30	thesesups.ups-tlse.fr Internet Source		<1 %

Exclude quotes	Off	Exclude matches	< 14 words
Exclude bibliography	Off		