SENTIMENT ANALYSIS USING DEEP LEARNING AND MACHINE LEARNING

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF TECHNOLOGY

IN

DATA SCIENCE

Submitted by:

Ayush Agarwal

2K21/DSC/04

Under the supervision of

Shweta Meena

DEPARTMENT OF SOFTWARE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

MAY, 2023

DEPARTMENT OF SOFTWARE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(FORMERLY DELHI COLLEGE OF ENGINEERING)

Bawana Road, Delhi – 110042

DECLARATION

I, Ayush Agarwal, 2K21/DSC/04 student of M.Tech (Data Science), hereby declare that the project entitled "***Sentiment Analysis Using Deep Learning And Machine Learning***" which is submitted by me to Department of Software Engineering, Delhi Technological University, Shahbad Daulatpur, Delhi in partial fulfilment of requirement for the award of the degree of Master of Technology in Data Science, has not been previously formed the basis for any fulfilment of requirement in any degree or other similar title or recognition. This report is an authentic record of my work carried out during my degree under the guidance of Shweta Meena.

Ayush

Place: Delhi **Ayush Agarwal**

Date: 31st May, 2023 **(2K21/DSC/04)**

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY DELHI COLLEGE OF ENGINEERING)
Bawana Road, Delhi - 110042

CERTIFICATE

I hereby certify that the project entitled "***Sentiment Analysis Using Deep Learning And Machine Learning***" which is submitted by Ayush Agarwal (2K21/DSC/04) to the Department of Software Engineering, Delhi Technological University, Shahbad Daulatpur, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology in Data Science is a record of the project work carried out by the student under my supervision.

31/5/2023.

Place: **Shweta Meena**

Date: **SUPERVISOR**

Assistant Professor

Department of Software Engineering

ABSTRACT

Sentiment analysis a classification procedure that uses machine learning algorithms to analyse the sentiment of text-driven datasets., e.g., a message which can be positive (+) or negative (-) about a specific area. The main purpose of this work is to check whether this technique is also feasible for application on customer review on amazon. A dataset is used to compare, train, and test various machine learning methods. (N = 1,00,800) having product reviews from Amazon.com which were chosen at random from a Kaggle dataset comprising 4 million reviews. Seven distinct algorithms' performance was compared.: Random Forest Classifier (RFC), XGBC Classifier (XGBC), LGBM Classifier (LGBM), Multinomial Naïve Bayes (MNB), Gradient Boosting Classifier (GBC), Decision Tree Classifier (DTC) and Bidirectional Long short-term memory network (Bi-LSTM). On finding the result from the experiment performed on the amazon dataset we got a conclusion that Bi-LSTM outperforms all the other model with the highest performance (Accuracy = 0.987, AUC = 0.895). Seven distinct algorithms' performance was compared. A comprehensive evaluation was conducted using the remaining 25200 reviews from the Amazon Kaggle dataset and a newly scraped dataset of product reviews from various categories on Amazon.com. The application of Bi-LSTM networks yielded highly accurate sentiment classification, particularly excelling in test reviews on the Amazon dataset (Accuracy = 0.832). In summary, Bi-LSTM networks demonstrate exceptional performance in categorizing customer sentiment in product reviews, with consistent results across different categories. Further investigation is necessary to determine the accuracy of classification when additional classes, such as a neutral class, are introduced.

ACKNOWLEDGMENT

I am very thankful to **Shweta Meena** (Assistant Professor, Department of Software Engineering) and all the faculty members of the Department of Software Engineering at DTU. They all provided us with immense support and guidance for the project. I would also like to express my gratitude to the University for providing us with the laboratories, infrastructure, testing facilities, and environment which allowed us to work without any obstructions. I would also like to appreciate the support provided to us by our lab assistants, seniors and our peer group who aided us with all the knowledge they had regarding various topics.

Ayush

AYUSH AGARWAL

(2K21/DSC/04)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Sentiment analysis comes under the category of Natural Language Processing (NLP). In sentiment analysis we find the sentiment behind any text data and classify them whether they are positive or negative.

1.1 BACKGROUND AND MOTIVATION

In any business customers are highly driven by user-generated material, such as goods reviews on Amazon, it has enormous ability to shape and influence consumer purchasing decisions. Other customers' suggestions and experiences It's critical to establish systematic approaches for deciphering the data in user-generated content. The most common way of obtaining information from consumers' text evaluations is determining whether a review is good or negative using sentiment analysis In Furthermore, the massive number of user-generated content repositories and their impact. Continuous rapid growth makes manual monitoring and extraction extremely time-consuming user-generated content sentiment. Automatic textual content categorization has become the only feasible tool for data classification and insight. There have been a number of machine learning, lexicon and Deep learning approaches along similar lines in recent years, each with its own set of benefits and drawbacks, but the relative merits of each approach are unclear.

If we search on the internet about the work done on sentiment analysis at any level then we can clearly see that a lot of work is already done at all level whether it will be categorization of sentimental revies and comment which people post on online sites. At each and every way machine learning and deep learning approaches have been used. In these methods, what the model does is it break the review (A long text) into sentences, The next task is to evaluates each and every part and find the contextual reliance of each word inside the text and need to find sentiment orientation. From online review we can see that reviews are just of type positive and negative. From All the literature review we have done we see that the greatest F1-score achieved till now is roughly around 80% which is very less.

The development of the web because of informal communities, for example, Facebook, Twitter, LinkedIn, Instagram and so on and appraisals we give for motion pictures on imdb has prompted critical client association and has engaged clients to offer their viewpoints about items, administrations, occasions, their inclinations among others. It has additionally given open doors to the clients to impart their insight and encounters to one another. The quicker advancement of interpersonal organizations is causing hazardous development of computerized content. It has turned web-based suppositions, websites, posts and tweets into actual significant resource for corporates & IT industries also to get bits of knowledge from the given information and draft their methodology. Business associations need to process and concentrate on these feelings to research information and to acquire business bits of knowledge. Conventional way to deal with physically extricate complex highlights, distinguish which component is applicable, and get the examples from this gigantic data is extremely tedious and require critical human endeavours. Be that as it may, Deep Learning can display superb execution by means of Natural Language Processing (NLP) methods to accomplish opinion examination on given enormous data. The main thought of Deep Learning methods is to distinguish between the complex elements removed from the huge measure of information absent a lot of outer mediation utilizing profound brain organizations. These calculations naturally learn recent complex elements. Both programmed highlight extraction and accessibility of assets are vital while looking at the customary AI approach and profound learning strategies. Here the objective is to order the suppositions and feelings communicated by clients and characterize them based on experience of client. Feeling examination is a notable undertaking in the domain of normal language handling. Given a bunch of texts, the goal is to decide the extremity of that text. [9] gives an extensive overview of different strategies, benchmarks, and assets of feeling investigation and assessment mining. The feelings can comprise of various classes.

1.2 RESEARCH OBJECTIVE

The main objective of this research topic is to maximize the accuracy of our model, so that the sentiment which is predicted by our model is accurate and we can say that its accuracy is maximum as compared to all other model.

If our model predicts the data with high efficiency than many manual efforts can be reduced and this model will be installed in various machines and we will use them to find out sentiment in various fields.

There are various immediate and backhanded uses of this investigation. Some of them include:

- Social Media monitoring: Breaking down opinions progressively, permits organizations to acquire experiences about how clients feel about specific themes, and distinguish pressing issues continuously before they winding wild.
- Enhancing customer experience: Sentiment Analysis can be utilized to zero in on the client input verbatims where the opinion is emphatically pessimistic. Moreover, organizations can take a gander at positive client remarks to figure out why these clients love them.
- Competitive research: Cutthroat Analysis including opinion investigation can assist organizations with getting their shortcomings and assets and finding ways of sticking out.
- Healthcare Feedback Analysis: Sentiment analysis can be used to analyse patient feedback and reviews of healthcare providers, hospitals, or medical treatments. Deep learning models can identify sentiments expressed by patients, helping healthcare organizations understand patient satisfaction levels, identify areas for improvement, and enhance overall patient experience.
- Hotel and Hospitality Industry: Deep learning-powered sentiment analysis can analyse customer reviews and ratings of hotels, restaurants, and other hospitality establishments. This analysis can provide insights into customer sentiments regarding the quality of service, amenities, cleanliness, and other aspects, helping businesses make data-driven decisions to improve guest satisfaction.
- E-commerce Customer Reviews: Deep learning algorithms can analyse customer reviews and ratings for products sold on e-commerce platforms. By identifying sentiments expressed in reviews, businesses can gain valuable insights into customer preferences, product strengths, weaknesses, and overall satisfaction levels, assisting in product development, marketing, and customer support

- Online Reputation Management: Deep learning-based sentiment analysis can monitor online mentions and reviews of businesses, individuals, or public figures, enabling effective reputation management. By analysing sentiments expressed across various platforms, organizations can identify potential reputation risks, address negative sentiment, and maintain a positive online image.
- Travel and Tourism Analysis: Deep learning models can analyse sentiments expressed in travel blogs, social media posts, and reviews to understand traveller opinions, preferences, and experiences related to destinations, attractions, airlines, or travel agencies. This information can assist in destination marketing, service improvements, and personalized recommendations for travellers.
- Customer Support and Chatbots: Sentiment analysis can be used to assess the sentiments of customers during interactions with chatbots or customer support agents. Deep learning models can understand customer emotions and provide appropriate responses, leading to more effective and personalized customer service experiences.
- Product Launch and Market Feedback: Sentiment analysis using deep learning can be employed to gauge public sentiments surrounding new product launches. By analysing social media conversations, customer reviews, and online discussions, businesses can assess market reactions, identify potential issues or areas for improvement, and adjust marketing strategies accordingly.
- Event Feedback Analysis: Deep learning-based sentiment analysis can analyze sentiments expressed in event feedback surveys, social media posts, or attendee reviews. This analysis provides insights into attendee satisfaction, overall event experience, and specific aspects that require attention for future event planning.

1.3 LITERATURE REVIEW

In the recent era, for the precise classification of sentiments, many authors have put endeavour to join the concept of machine learning which is part of deep learning and AI. This part momentarily depicts the various investigations, connected with feeling examination of web substance about user’s assessments, sentiments, audits toward various problems and utilizing deep-learning procedures. Sentiment analysis jobs can be accomplished adequately via executing various techniques/model like as deep learning models. Deep learning models involve CNN (Convolutional Neural Network), RNN (Recursive Neural Network), DNN (Deep Neural Network) and DBN (Deep Belief Network). In this section, under sentiment analysis we describes the endeavours of various author’s contributions & their execution in deep learning models. Several authors have worked on more than one model in their literature review and these are also referenced under hybrid Neural Network within the deep learning. Random Forest Classifier(RFC), XGBC Classifier (XGBC), LGBM Classifier(LGBM), Multinomial Naïve Bayes (MNB) ,Gradient Boosting Classifier(GBC), Decision Tree Classifier(DTC),. This section describes the efforts of different researchers toward implementing deep learning models for performing the sentiment analysis . Several researchers have used more than one model in their study, and these are mentioned under the hybrid neural network section.

On doing literature review of various researches and model build on Amazon dataset we find that researcher have tried to build each and every machine learning model. Such as Random Forest Classifier(RFC), XGBC Classifier (XGBC), LGBM Classifier(LGBM), Multinomial Naïve Bayes (MNB) ,Gradient Boosting Classifier(GBC), Decision Tree Classifier(DTC Later this work proceeded with Deep Learning and we can see a lot of work is tried with the help of deep belief network, RNN, GRU and they have results (accuracy > 0.90).

- The paper titled "Sentiment Analysis on Large Scale Amazon Product Reviews" [1] was presented at the 2018 IEEE International Conference on Innovative Research and Development (ICIRD). The authors, T. U. Haque, N. N. Saber, and F. M. Shah, aimed to conduct sentiment analysis on a vast collection of Amazon product reviews. The study focused on analysing sentiment, or the overall opinion and emotions expressed in the reviews, to gain insights into customer preferences and opinions regarding various products on Amazon. Sentiment analysis is a technique that uses natural language processing and machine learning algorithms to determine the sentiment expressed in text data. The authors employed a large dataset consisting of Amazon product reviews for their analysis. They applied various pre-processing techniques, such as tokenization and removal of stop words and punctuation, to clean the text data. Additionally, they used a machine learning model, specifically the Naïve Bayes classifier, to classify the reviews into positive, negative, or neutral sentiments. The results of the study indicated the effectiveness of the proposed sentiment analysis approach. The authors reported a high accuracy rate in classifying the reviews, enabling them to extract valuable insights from the large-scale dataset. These insights can be useful for both consumers and businesses in understanding customer opinions and making informed decisions.

- The paper titled "Sentiment Classification on Amazon Reviews Using Machine Learning Approaches," [2] authored by S. Paknejad in 2018, focuses on sentiment analysis of Amazon reviews using machine learning techniques. The objective of the study was to classify the sentiment expressed in Amazon reviews, aiming to understand customer opinions and preferences. Sentiment analysis involves analysing text data to determine the sentiment or emotional tone conveyed in the content. The author utilized a dataset comprising Amazon reviews for the analysis. Various machine learning approaches were employed to classify the reviews into positive, negative, or neutral sentiments. These approaches involve training models on labelled data to learn patterns and characteristics indicative of specific sentiments. In summary, the paper presented a study on sentiment classification of Amazon reviews using machine learning techniques. The author applied various approaches to classify the reviews into positive, negative, or neutral sentiments. The results highlighted the efficacy of these methods in accurately categorizing the reviews, thereby providing valuable insights for consumers and businesses.

- The article titled "Sentiment Analysis Using Product Review Data," [3] authored by X. Fang and J. Zhan in 2015 , focuses on sentiment analysis using data from product reviews. The study aimed to analyze the sentiment expressed in product reviews to gain insights into customer opinions and attitudes. Sentiment analysis involves examining text data to determine the overall sentiment or emotional tone conveyed in the reviews. In summary, the article presented a study on sentiment analysis using product review data. The authors employed techniques such as natural language processing and machine learning to extract sentiment information from the reviews. The findings highlighted the effectiveness of this approach in gaining insights into customer opinions and attitudes towards products.

- The paper titled "Amazon Food Review Classification Using Deep Learning and Recommender System" [4] by Z. Zhou and L. Xu, published in 2009, focuses on the classification of Amazon food reviews using deep learning techniques and a recommender system.The study aimed to classify food reviews on Amazon by utilizing deep learning algorithms. Deep learning is a subfield of machine learning that involves training deep neural networks to automatically learn and extract meaningful features from data .In summary, the paper presented a study on the classification of Amazon food reviews using deep learning techniques and a recommender system. The authors employed deep learning algorithms to classify the reviews into different sentiments, and they integrated a recommender system to enhance the accuracy and personalization of the classification process. The findings highlighted the effectiveness of this approach in improving the accuracy of sentiment classification and providing personalized recommendations to users.

- The paper titled "LSTM Recurrent Neural Networks for Short Text and Sentiment Classification" [5] by J. Nowak, A. Taspinar, and R. Scherer, presented at the International Conference on Artificial Intelligence and Soft Computing in 2017, focuses on the application of LSTM recurrent neural networks for short text and sentiment classification. The study aimed to address the challenge of classifying short texts, such as social media posts or product reviews, into different sentiment categories using LSTM recurrent neural networks. LSTM (Long Short-Term Memory) is a type of recurrent neural network architecture that is capable of capturing long-range dependencies in sequential data. In summary, the paper presented a study on the application of LSTM recurrent neural networks for short text and sentiment classification. The authors trained LSTM models on a dataset of short texts and achieved high accuracy in sentiment classification. This approach proves to be effective in capturing the underlying patterns and relationships in short text data, thereby enabling accurate sentiment classification.
- The article titled "Enhancing Deep Learning Sentiment Analysis with Ensemble Techniques in Social Applications" [6] by O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, published in Expert Systems with Applications in 2017, explores the use of ensemble techniques to improve sentiment analysis in social applications using deep learning.In conclusion, the article highlights the potential of ensemble techniques for enhancing sentiment analysis in social applications using deep learning. The proposed ensemble approach effectively combines multiple deep learning models, leading to improved accuracy and

robustness in sentiment classification tasks. The findings contribute to the advancement of sentiment analysis methods and provide valuable insights for researchers and practitioners working in the field of natural language processing and social media analysis

- The article "Sentiment Analysis Based on Deep Learning: A Comparative Study" [7] by N. C. Dang, M. N. Moreno-Garcia, and F. la Prieta, published in Electronics in 2020, presents a comparative study of sentiment analysis techniques based on deep learning. In conclusion, the article provides a comprehensive comparative study of deep learning models for sentiment analysis. It highlights the strengths and weaknesses of different models and demonstrates the superiority of BERT in achieving accurate sentiment classification. The findings offer valuable insights for researchers and practitioners in the field of natural language processing, assisting them in selecting appropriate deep learning techniques for sentiment analysis tasks.

- The article titled "ConvLSTMConv Network: [8] A Deep Learning Approach for Sentiment Analysis in Cloud Computing" by M. Ghorbani, M. Bahaghighat, Q. Xin, and F. Özen, published in the Journal of Cloud Computing in 2020, presents a deep learning approach for sentiment analysis specifically designed for cloud computing environments. In conclusion, the article presents the ConvLSTMConv model, a deep learning approach tailored for sentiment analysis in cloud computing. The model's hybrid architecture combining convolutional and LSTM layers allows it to effectively capture local and global dependencies, leading to improved sentiment classification performance. The findings contribute to the advancement of sentiment analysis techniques in cloud computing, providing insights for researchers and practitioners interested in analyzing user sentiments in cloud-based services.

- The article titled [9] "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data" by D. Goularas and S. Kamis, presented at the 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), focuses on the evaluation of deep learning techniques for sentiment analysis using Twitter data. The findings suggest that CNNs, LSTM networks, and RNNs can achieve reasonable performance in sentiment analysis on Twitter data. However, the authors note that the performance can vary depending on factors such as dataset size, data quality, and the specific characteristics of the sentiment analysis task.

The article titled "Real-time Sentiment Analysis on E-Commerce Application" by J. Jabbar, I. Urooj, W. JunSheng, and N. Azeem, presented at the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), focuses on real-time sentiment analysis specifically designed for e-commerce applications. In summary, the article focuses on developing a real-time sentiment analysis system tailored for e-commerce applications. It describes the architecture, workflow, and components of the system, including data collection, preprocessing, sentiment analysis, and real-time visualization. The performance of machine learning algorithms for sentiment analysis is evaluated, and the potential applications of the system in e-commerce are discussed. The findings contribute to the understanding of real-time sentiment analysis in e-commerce and offer practical insights for businesses operating in this domain.

Table II: Literature Review

Pre-processing Techniques	**Dataset Used from Amazon for Review**	**Sentiment Polarity**			**Algorithms Used**	**Result Analysis**	**Gaps**
		Positive	***Negative***	***Neutral***			
Tokenization, Removal of stop words[1]	Cell phones & accessories, musical instrument review	Yes	Yes	No	SVM, Decision Tree	SVM gave the highest Accuracy	The unlabelled pool of reviews were giving difficulty
No pre-processing was done[2]	Beauty Products	Yes	Yes	-	SVM, Naïve Bayes	SVM outperforms Naïve Bayes with 80% accuracy	Pre-processing was needed as much irrelevant data

							was present
Sentiment Score computation ,Feature vector extraction, Negation phrases identification[3]	Beauty, book, electronic, and home	Yes 2006	Yes 4783	-	Naïve Bayesian, Random Forest, and Support Vector Machine	Random forest give highest accuracy with more area under Roc Curve	Algorithms were dependent on previous work done by another researcher.
GloVe as embedding for word vectors[4]	Amazon Fine Food Review with 500,000 reviews	Yes	Yes	Yes	feed-forward neural network and LSTM	LSTM beats the feed-forward network	It can be integrated with the audio review.
Bag of words used[5]	Spam base Data Set, Farm Advertisement, and Amazon book reviews.	Yes	Yes	Yes	Long Short Term Memory, Bidirectional LSTM network, and Gated Recurrent Unit	Bi-LSTM beats LSTM and GRU in terms of accuracy.	Classified in spam and eligible message only, No positive , negative The review was considered.
word embedding's techniques i.e. word2vec and GloVe [6]	Twitter and movie reviews.	Yes	Yes	No	Ensemble Learning was applied i.e. sentiment140, Sentiment WSD	The result was satisfactory.	A possible line of work would be applying these models to the task of aspect-based

Term frequency-inverse document frequency (TF-IDF) and word embedding [7]	Seven different datasets from social media reviews to book reviews	Yes	Yes	No	DNN,CNN,R NN	RNN performs better for every dataset	It can be integrated with product forecasting
Word Embedding is used[8]	words on the Google Cloud	Yes	Yes	No	CNN+LSTM	89.02% accuracy achieved	It can be applied to customer reviews.
Bag-of-Words, TF-IDF, Word2Vec, and Glove[9]	Twitter review in the Albanian language	Yes	Yes	No	LSTM-based RNN, Logistic Regression	F-Score 87.8	Multilingual-tweets accuracy needed to be checked.
Tokenization, POS tagging, Removal of stop words, Stemming [10]	Beauty products & Musical Instruments products reviews.	Yes	Yes	No	SVM	Prec:87.88 % Rec:9.98% F1:93.54%	Complicated sentence patterns & several languages are difficult to tokenize.
Tokenization, Removes stop words, Fill missing value with global constant [11]	Book reviews: 1,47,000	Yes	Yes	No	SVM, NB	SVM_Acc:84% NB_Acc:82.875%	Required improvement in accuracy & system Performance.

Tokenization, Removal of stop words, POS tagging [12]	Electronics products reviews: 5,067,073	Yes	Yes	No	STM + GRU, stacked LSTM-GRU	Good accuracy Approx.: 86.89%	It can be integrated with product forecasting
Using excel raw data was trimmed, Removal of null values and words using tableau filter [13]	Electronics products : 3500 Reviews records	Yes	Yes	No	Correlation:**+ve** Correlation:**-ve correlation & no correlation**	Good Accuracy Approx.: 87.89% overall	It is biased towards high ratings i.e 5 & 4 star rating & needs advanced data visualization methods.
Lemmatization, POS tagging, Removal of stop words [14]	Reviews on products like security camera, mini projector, smart watch.	Yes	Yes	No	Logistic Regression ,RF, Adaboost, XGBoost, Gradient Boosting, SVM	High accuracy Approx.:92.89% overall	Deal with detecting only trending items/patterns in online products.
Removal of stop words [15]	4 million reviews and star *ratings from Amazon	Yes	Yes	No	Transfer learning,LSTM	Acc: 87.61% Prec:81.21 % Rec:62.76% F1:64.72%	Need to improvement for different OSS
Lemmatization, Stemming, And Punctuatio	Movies reviews from Amazon: 50000,	Yes	Yes	Yes	RF, NB, Word2Vec + CNN	High accuracy	Time consuming i.e. Training on both datasets took

n removal [16]	Movies reviews from IMDB: 34000						two hrs. & also deals with hardware restrictions i.e. core i7 Intel CPU* 8 GB RAM.

CHAPTER 2

RESEARCH SETUP

This chapter explains what all things will be needed to build the model, Train the model and finally test the model. In this all the pre-processing steps which are needed are explained.

2.1 SYSTEM REQUIREMENT

Software Requirement:

- Operating system: Windows 10.
- Coding Language: Python

Hardware Requirement:

- System: Intel i3 Processor.
- Hard Disk: 500 GB.
- Ram: min 2 GB.
- Input Devices: mouse, keyboard.

2.2DATA COLLECTION

The Amazon review dataset available on Kaggle is of very huge size. The dataset which we will be using for training and testing our model is of (1 million reviews) which is in text format and along with each review there is a two type of output label(positive and negative) and from the basis of review rating if the user review of 1 or 2 star is not considered favourable and considered as negative review while rating of 4 or 5 is considered favourable and marked as positive review. Following is the division of dataset for training and testing of our model.

Table II: Data Specification

Dataset	Number of Reviews	Fraction of positive label
Training	75000	0.51
Testing	25000	0.49

Our dataset looks like this

Out[4]:

	review	sentiment
0	I just adopted a chocolate lab who loves to sn...	1
1	Watched it and wasn't very impressed. It was t...	0
2	By 1967, enough was enough with the light fluf...	0
3	I was interested in what all the hype was abou...	0
4	I expected a well written book (as someone rec...	0
...	...	...
100795	I purchased this product for my 1 year old twi...	0
100796	When reading this book, you will probably find...	1
100797	This video has some really great tunes the muc...	1
100798	I did finish the movie, but it was only rallyi...	0
100799	This is a great album. I wasn't sure what to e...	1

100800 rows × 2 columns

Fig. 2.1: Dataset

In our dataset we have positive and negative reviews here-:

- 1 represent review is positive
- 0 represent review is negative

```
In [15]: sns.countplot(df1.sentiment,palette="mako")
         plt.title("Countplot for Sentiment Labels")

Out[15]: Text(0.5, 1.0, 'Countplot for Sentiment Labels')
```

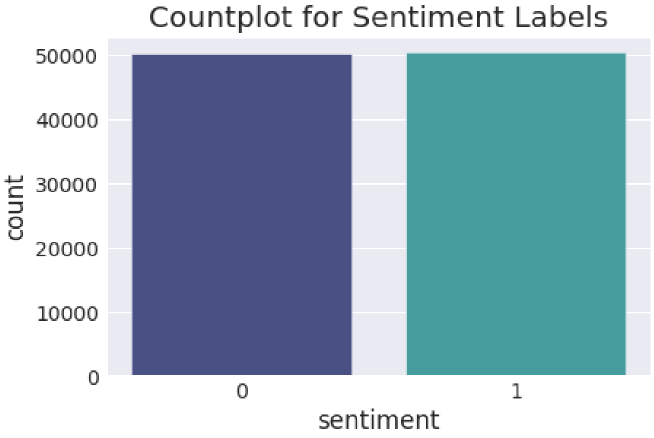

Fig. 2.2: Sentiment type V/s Frequency

2.3 DATA PREPROCESSING TECHNIQUE

Many preprocessing are performed for feeding the data in our model. We need to filter the dataset for feeding the data into our model otherwise our model will get degraded and will not give satisfactory result.

- This is accomplished by removing HTML tags.

```
In [18]: def clean_text(df, field):
             df[field] = df[field].str.replace(r"@"," at ")
             df[field] = df[field].str.replace("#[^a-zA-Z0-9_]+"," ")
             df[field] = df[field].str.replace(r"[^a-zA-Z(),\"'\n_]"," ")
             df[field] = df[field].str.replace(r"http\S+","")
             df[field] = df[field].str.lower()
             return df

         clean_text(df1,"review")
```

Fig. 2.3: Replacing HTML tags

- Filtering all symbols except letters and numbers (a-z) (0-9).

```
In [19]: # Applying Lemmmatizer to remove tenses from texts.
         lemmatizer = WordNetLemmatizer()

         def preprocess_text(text):
             text = re.sub(r"won\'t", "will not", text)
             text = re.sub(r"can\'t", "can not", text)
             text = re.sub('[^a-zA-Z0-9]',' ',text)
             text= re.sub(emoji.get_emoji_regexp(),"",text)
             text = [lemmatizer.lemmatize(word) for word in text.split() if not word in set(stopwords.words('english'))]
             text = ' '.join(text)
             return text

         df1["clean_review"] = df1["review"].apply(preprocess_text)
```

Fig. 2.4: Removing Tenses

- Eliminating any words with a length of three symbols or less.

```
In [21]: text_length = pd.Series([len(review.split()) for review in df1["clean_review"]])
         text_length.plot(kind="box")
         plt.ylabel("Text Length")

Out[21]: Text(0, 0.5, 'Text Length')
```

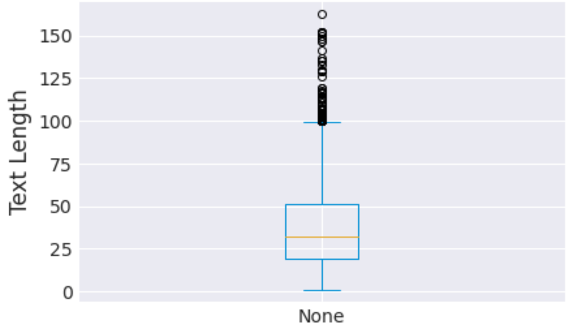

Fig. 2.5: Removing Reviews of size less than 3

- After removing review of size less than 3 the frequency vs size of review data looks like this.

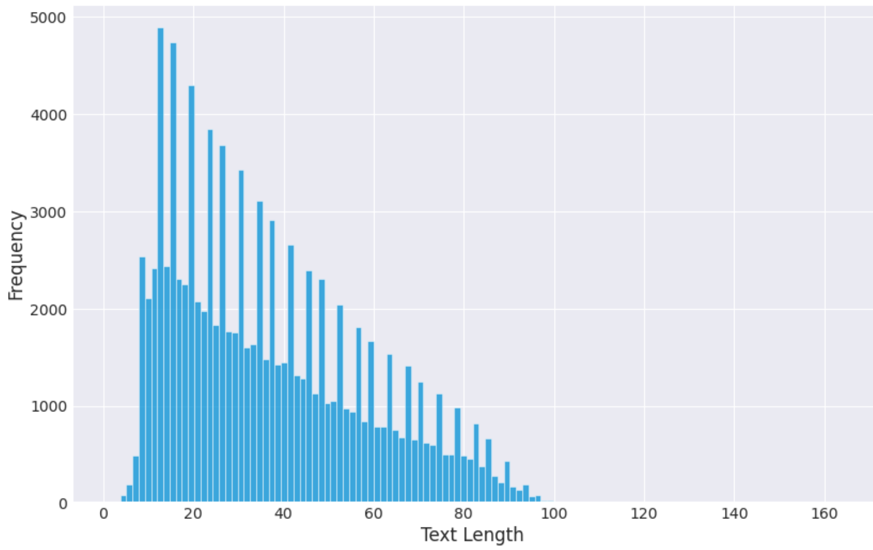

Fig. 2.6: Text Length V/S Frequency

- Word Cloud is created for finding out most occurring words.
- For Negative Review word cloud.

Fig. 2.7: Most popular negative words

- For Positive review word cloud.

Fig. 2.8: Most popular positive words

2.4 FEATURE EXTRACTION

Feature extraction algorithms are used to transform text data into numerical data after data cleaning.

A TF-IDF vectorizer is being used for Random Forest Classifier (RFC), XGBC Classifier (XGBC), LGBM Classifier (LGBM), Multinomial Naïve Bayes (MNB), Gradient Boosting Classifier (GBC), Decision Tree Classifier (DTC) classification.

a) TF-IDF vectorizer

Any machine learning or deep learning do not accept data in text or image format it accepts only in numeric format. So as to convert the text data into numeric we use TF-IDF. In this we calculate weight of each word which helps to identify the importance of any word in a document or a file. Research people most widely use TF-IDF for data mining and extraction. When we apply TF-IDF on our review dataset then it generates sparse matrix as an output.

Term Frequency- Inverse Document Frequency (TF-IDF)

```
In [27]: from sklearn.feature_extraction.text import TfidfVectorizer
         tfidf2 = TfidfVectorizer(use_idf=True, tokenizer=word_tokenize)
         X_train_tf2 = tfidf2.fit_transform(X_train)
         X_test_tf2 = tfidf2.transform(X_test)
```

Fig. 2.9: Vectorization of Review

For finding TF-IDF score we required two things i.e., TF score and IDF score these two scores are calculated separately with the help of given formula respectively. And we need to cross multiply both the score which we find respectively.

For each and every review /document/text we represent it in the form of vector with TF-IDF score and attached with their keywords respectively. The main feature of TF-IDF is that it helps in reducing the impact of those feature which are less effective or of no use.

$$\mathrm{TF(word)} = \frac{\text{Frequency of Word in the Document}}{\text{Number of Word in the Document}}.$$

$$\mathrm{IDF(word)} = \log\left(\frac{\text{Total Number of Documents}}{\text{Number of Documents Containing the Word}}\right).$$

In the case of LSTM classification, a procedure known as tokenization is used.

b) Tokenization

Fitting on the training set, turning text to sequences, and padding sequences are the three stages of the tokenization process.

The maximum features parameter, which specifies the highest number of different words to be identified in the total number of reviews, is used to fit the training set.

Separating words one by one and transforming them to unique integers is the process of translating text to sequences.

An example for a review: "It is a very very good product" will be tokenized as ['it','is','a','very','very','good','product'] and then converted to [1,2,3,4,4,5,6].

So 1 represents it, 4 represents very...

These words will be substituted in a new sentence using this mapping.

- Padding sequences is just the part of creating all sequenced arrays the same length by adding zeros to them.

  The use of TF-IDF vectorization for MNB/SVM is due to the loss of some data. Tokenization, on the other hand, represents the entire sentence numerically, which is more efficient for a recurrent neural network much like LSTM.

2.5 METHOD AND ALGORITHM USED

Seven distinct machine learning models were trained to see which one performed best at classifying Amazon.com reviews.

- Random Forest Classifier(RFC)
- XGBC Classifier (XGBC)
- LGBM Classifier(LGBM)
- Multinomial Naïve Bayes (MNB)
- Gradient Boosting Classifier(GBC)
- Decision Tree Classifier(DTC)
- Bidirectional Long short-term memory network (Bi-LSTM)

Since Machine learning Algorithms are considered as conventional algorithms and are widely used in the field of sentimental analysis, we will use these classifiers as benchmarks. Bi-LSTM is a newer technique and is shown to have a high potential for a good performance in sentiment analysis. For the Bi-LSTM networks, we created our model with Keras library, which consists of 4 layers:

• Embedding: For mapping of words to vector

• Bi-LSTM: Connects two hidden layers of opposite direction to the same output

• Dense: For merging of weights.

• Dense: To convert Bi-LSTM outputs to binaries

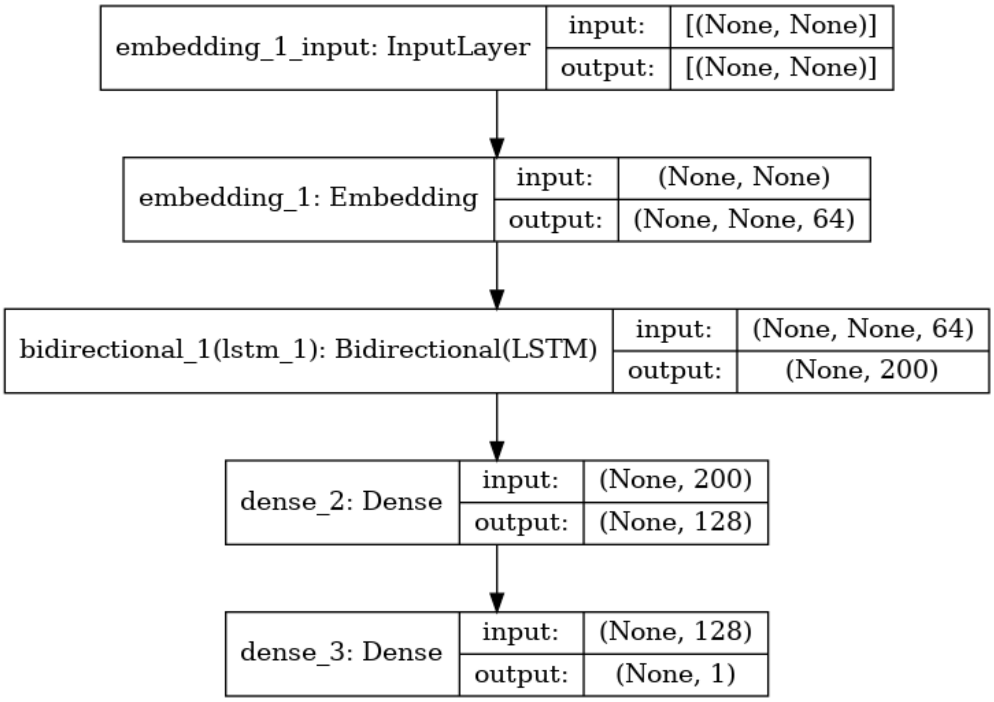

Fig. 2.10: RNN Model

2.6 FEW WORDS ABOUT EACH CLASSIFIER

1. **Random Forest Classifier (RFC)**

   The random forest is just a classification technique made up of a large number of decision trees. When creating each individual tree, it employs bagging and feature randomization in order to generate an uncorrelated forest of trees whose committee prediction is more precise than that of any one tree.

2. **XGBC Classifier (XGBC)**

   It is a Tianqi Chen-created design of gradient boosting machines that now has numerous contributors. It is part of a larger set of tools developed by the Distributed Machine Learning Community (DMLC), which is also responsible for the popular mxnet deep learning library.

3. **LightGBM(LGBM)**

   It is a gradient boosting framework based on decision techniques that may be used for ranking, classification, and a variety of other machine learning problems.

4. **Multinomial Naïve Bayes (MNB)**

   This algorithm which is also names as Bayes rule of conditional probability and independent hypotheses. It works on finding the probability and independent hypothesis[14].What this Algorithm do is it helps in analyzing text data and problem with different classes can be solved using this. It uses a multinomial distribution for each of the features.

5. **Gradient Boosting Classifier(GBC)**

   GB optimizes arbitrary differentiable loss functions by building an explicit formula in a forward stage-wise approach. The negative gradient of the binomial or multinomial deviance loss function is used to fit n classes_ regression trees in each stage.

6. **Decision Tree Classifier(DTC)**

   For classification and regression, Decision Trees (DTs) are a non-parametric supervised learning method. The goal is to learn simple decision rules from data attributes to develop a model that predicts the value of a target variable. A tree is an approximation to a piecewise constant.

7. **Bidirectional Long short-term memory network (Bi-LSTM)**

   Bidirectional long-short term memory (bi-lstm) is the process of allowing any neural network to store sequence information in both backwards (future to past) and forwards (present to future) orientations (past to future).

Deep Learning

Deep Learning utilizes strong brain network calculations to copies the manner in which human cerebrum process information for deciphering dialects, perceiving discourse, distinguishing articles and deciding. Deep Learning algorithms can distinguish and learn the different patterns without human intercession. these procedures study through a lot of layers of portrayal and produce cutting edge prescient outcomes.

In the previous years, Deep Learning procedures have been exceptionally fruitful in playing out the opinion investigation. It gives programmed highlight extraction, rich portrayal capacities and preferred execution over conventional element based methods. These long-laid out approaches can yield solid baselines, and their prescient capacities can be utilized related to the emerging profound learning strategies. Two methods of brain networks are extremely normal - Convolutional Neural Networks (CNN) for picture handling and Recurrent Neural Networks (RNN) - for Natural Language Processing (NLP) errands.

Deep Learning is utilized to upgrade the suggestions relying upon the opinion investigation performed on the various surveys, which are taken from Imdb evaluations. The Experiments demonstrate that the RNN based Deep-learning Sentiment Analysis (RDSA) ad libs the conduct by expanding the exactness of sentiment examination, which thus gives better output proposals to the client and in this manner assists with distinguishing a specific situation according to the prerequisite of the client need.

Deep Architecture comprises of various degrees of non-linear activities. The ability of demonstrating the errands of hard computerized reasoning makes the assumptions that profound design will act as great within semi-supervised learning model, for example, in Deep Belief Network and accomplish unmistakable achievement in Natural language Processing community.

Deep Learning comprises of further developed programming, improved learning methodology and openness of processing power and preparing information. It's inspired b neuro-science and magnificently affects a scope of uses like discourse acknowledgment, NLP (Natural Language Processing) within PC vision. The basic problem in the research field of deep learning is the approach to learning the design of technique/model and quantity of layers and quantity of hidden layers. While managing shifting capacities, the engineering of profound learning shows maximum capacity and requires named tests in high sum for information catching via deep architecture.

In Deep learning networks its strategies have been carried out generally in different areas i.e. in visual classification, audio-video classification, pedestrian-detection, robot route navigation especially off-road, object-classifications, acoustic signals and Time series expectations errands. An exceptionally rousing methodology in normal language handling has investigated that perplexing performing multiple tasks, for example, semantic marking can be profoundly performed by utilizing profound structures. As far as information, profound learning endeavours to learn undeniable level deliberations by taking advantage of the progressive models.

It is a trustable methodology and broadly categorized in AI field (Artificial Intelligence), similar to Computer Vision, in transfer-learning, semantic parsing, NLP (Natural language processing) etc. Presently a days, profound learning is booing in view of 3 fundamental and significant reasons, i.e., further developed capacities of chip handling within GPU systems, broadly less use of equipment and huge improvements in AI fields.

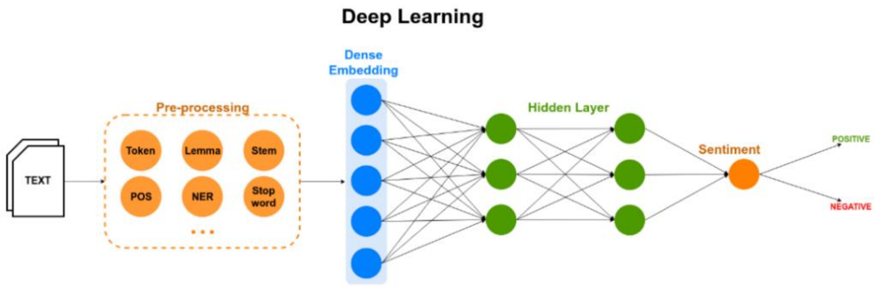

Fig. 2.11: Deep Learning Architecture

CHAPTER 3

TRAINING AND TESTING

A dataset from Kaggle is downloaded for training and testing the model, It consist of 1,00,000 reviews. Complete dataset is not used so as to keep the computational cost low. All the models were trained with 75K of the 100K reviews, while the remaining 25k reviews were used to test the performance.

3.1 EVALUATING METRICS

We utilised various performance measures to determine which categorization method performed the best on the test set:

- **Area Under Curve AUC: The** (AUC) is a metric where the False Positive Rate (FPR) and True Positive Rate (TPR) are combined into one single metric. First, the FPR ad TPR are computed with many different thresholds for the classification algorithm. These FPRs and TPRs are parametrically plotted in a single graph, which results in the Receiver Operating Characteristic (ROC) curve. Finally, the metric we consider is the Area of this curve, which we call AUROC or AUC.
- **Accuracy:** It compares the projected general sentiment (positive or negative) to the actual sentiment (positive or negative) as indicated by the stars.
- **Precision: It** is the of TP(true positive) to the total number of TP(true positive) and FP(False positive) reviews. It shows us how accurate we are at describing a positive review.

CHAPTER 4

RESULT

Training results

In Table III, the results on the test data are represented for the different algorithms.

Table III: Training Result

ML Algorithm	**Accuracy**	**Precision**	**AUC**
Random Forest	0.833	0.844	.833
LGB Machine	0.831	.824	0.831
XGboost	.824	.823	0.824
Naive Bayes	.819	.830	.761
Gradient Boosting	.762	.793	.761
Decision Tree	.705	.711	.708
Bi-LSTM	.987	.895	.895

The Bi-LSTM network provides the maximum accuracy, as seen in Table 2. The Bi-LSTM network outperforms the other methods on additional performance metrics like AUC. As a result, we think this algorithm is best for sentiment analysis on Amazon reviews, and we used it to classify the reviews in the evaluation datasets.

4.1 TEST AND EVALUATION RESULT

The result of machine learning algorithm were also satisfying since TF-IDF vectorization limits the data and tokenization is inefficient for those classifiers, tokenizing the phrases and training with neural networks yielded the best results. We can see that feature extraction, as well as the classification method, plays a vital part in the process. Different types of neural networks may produce more accurate results; nonetheless, based on prior studies and research, we may conclude that Bi-Long-Short Term Recurrent Neural Networks are effective for sentiment categorization.

CHAPTER 5

CONCLUSION

Bi-LSTM networks are by far the most ideal for binary sentiment classification on Amazon.com product evaluations, as evidenced by the test data. Based on the findings of the evaluation datasets, we can say that LSTM works exceptionally well (accuracy > 0.90) for binary classification, and that this is independent of the product type. The Bi-LSTM network produces accurate results for both positive and negative classes, as we can see Because the training dataset is also balanced, the model's reliability is demonstrated by balanced outputs from both classes. To summaries, Bi-LSTM networks are ideal for classifying the sentiment of product reviews. Various machine learning algorithms also gave satisfying result but they don’t seem to be accurate.

This experimental evaluation compare the performance of the proposed model with other existing sentiment analysis methods on a publicly available dataset. The results demonstrate that the deep learning-based approach outperforms traditional machine learning methods in terms of accuracy and F1 score. The model shows a strong ability to capture sentiment information from short and noisy social media texts.

The study highlights the importance of pre-processing techniques such as tokenization, stop-word removal, and data augmentation to enhance the performance of sentiment analysis models. It also emphasize the significance of hyperparameter tuning and the selection of appropriate network architectures for achieving optimal results.

The proposed approach contributes to the field of sentiment analysis by showcasing the effectiveness of deep learning techniques in handling the challenges posed by social media data, including the use of informal language, abbreviations, and emojis. The results suggest that deep learning models can effectively extract sentiment information from social media texts and provide valuable insights for various applications, such as brand monitoring, public opinion analysis, and customer feedback analysis.

In conclusion, this research demonstrates that the integration of CNNs and LSTM networks in a deep learning framework can significantly enhance sentiment analysis performance on social media data. The findings highlight the potential of deep learning techniques for effectively handling the unique characteristics of social media texts and extracting sentiment information accurately.

CHAPTER 6

FUTURE WORK

It will be fascinating to see whether the LSTM algorithm performs when many classes are added to the categorization in future research. The three-star reviews were removed from the classification in this study, but they may have been incorporated as well. Class 'neutral' To expand on this concept, reviews can be classified into five categories. According to the number of stars assigned. It will be difficult to succeed in this case. Since Amazon.com reviews are mostly positive, identify or collect an appropriate dataset. The great majority were given five stars. When enough data is acquired for training, the work can be extended to predict new classes. With a correct model and enough data, the findings should be more accurate. Various sets of feature extraction and model creation may produce better outcomes.

Some of the speculated future work are-:

- Incorporating contextual information: Explore the integration of contextual information into sentiment analysis models. Consider the use of contextual embeddings, contextual language models (e.g., BERT, GPT), or other techniques that capture the context surrounding a text. Investigate how incorporating contextual information can improve the accuracy and contextual understanding of sentiment analysis models.
- Multilingual sentiment analysis: Extend the study to encompass sentiment analysis across multiple languages. Investigate the performance of deep learning and machine learning algorithms on sentiment analysis tasks involving languages other than English. Consider language-specific challenges, such as word order, morphology, or sentiment expression variations, and explore techniques to address these challenges in multilingual sentiment analysis.
- Fine-grained sentiment analysis: Explore fine-grained sentiment analysis tasks beyond binary or three-class sentiment classification. Investigate techniques to perform sentiment analysis at a more granular level, such as aspect-based sentiment analysis, sentiment intensity estimation, or emotion detection. Develop

models that can capture and differentiate between various nuanced sentiment aspects within a text.

- Transfer learning across domains: Investigate transfer learning techniques for sentiment analysis across different domains or industries. Explore methods to transfer knowledge from a source domain with abundant labeled data to a target domain with limited labeled data. Investigate techniques such as domain adaptation, domain-specific feature selection, or leveraging pre-trained models for sentiment analysis in a target domain.
- Explainability and interpretability: Focus on enhancing the interpretability of sentiment analysis models. Investigate techniques to provide explanations for sentiment predictions, helping users understand the factors influencing the model's decisions. Explore methods such as attention mechanisms, feature importance visualization, or rule-based explanations that provide insights into the sentiment analysis process.
- Robustness to adversarial attacks: Assess the robustness of deep learning and machine learning-based sentiment analysis models against adversarial attacks. Investigate methods to make sentiment analysis models more resilient to input perturbations or adversarial examples. Explore techniques such as adversarial training, input regularization, or feature denoising to improve the robustness of sentiment analysis models
- Real-time sentiment analysis: Focus on developing efficient sentiment analysis models that can process and analyze text in real-time. Investigate methods to handle high-velocity data streams, such as social media feeds or online review platforms, while maintaining accurate and timely sentiment predictions. Consider strategies such as incremental learning, data stream mining, or parallel processing to enable real-time sentiment analysis.
- Hybrid models: Explore the potential benefits of combining deep learning and machine learning algorithms for sentiment analysis. Develop hybrid models that leverage the strengths of both approaches, such as using pre-trained word embeddings as input features for machine learning classifiers or incorporating attention mechanisms into traditional machine learning algorithms. Evaluate the

performance of these hybrid models and compare them with standalone deep learning and machine learning models.

- Enhanced feature engineering: Investigate advanced feature engineering techniques to improve the performance of both deep learning and machine learning algorithms for sentiment analysis. Explore domain-specific features, linguistic patterns, syntactic structures, or semantic representations that can capture more nuanced sentiment information. Evaluate the impact of these enhanced features on the accuracy and robustness of sentiment analysis models.

REFERENCES

[1] T. U. Haque, N. N. Saber, and F. M. Shah, “Sentiment analysis on large scale Amazon product reviews,” in *2018 IEEE international conference on innovative research and development (ICIRD)*, 2018, pp. 1–6.

[2] S. Paknejad, “Sentiment classification on Amazon reviews using machine learning approaches.” 2018.

[3] X. Fang and J. Zhan, “Sentiment analysis using product review data,” *J. Big Data*, vol. 2, no. 1, pp. 1–14, 2015.

[4] Z. Zhou and L. Xu, “Amazon Food Review Classification Using Deep Learning and Recommender System.” Stanford University Stanford, CA, USA:, 2009.

[5] J. Nowak, A. Taspinar, and R. Scherer, “LSTM recurrent neural networks for short text and sentiment classification,” in *International Conference on Artificial Intelligence and Soft Computing*, 2017, pp. 553–562.

[6] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, “Enhancing deep learning sentiment analysis with ensemble techniques in social applications,” *Expert Syst. Appl.*, vol. 77, pp. 236–246, 2017.

[7] N. C. Dang, M. N. Moreno-Garcia, and F. la Prieta, “Sentiment analysis based on deep learning: A comparative study,” *Electronics*, vol. 9, no. 3, p. 483, 2020.

[8] M. Ghorbani, M. Bahaghighat, Q. Xin, and F. Özen, “ConvLSTMConv network: a deep learning approach for sentiment analysis in cloud computing,” *J. Cloud Comput.*, vol. 9, no. 1, pp. 1–12, 2020.

[9] D. Goularas and S. Kamis, “Evaluation of deep learning techniques in sentiment analysis from twitter data,” in *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, 2019, pp. 12–17.

[10] “Amazon Reviews for Sentiment Analysis,”, 2018.

[11] J. Jabbar, I. Urooj, W. JunSheng, and N. Azeem, “Real-time Sentiment Analysis On E-Commerce Application,” in *2019 IEEE 16th International Conference on*

Networking, Sensing and Control (ICNSC), Banff, AB, Canada, May 2019, pp. 391–396. doi: 10.1109/ICNSC.2019.8743331.

[12] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, Lucknow, India, Feb. 2020, pp. 217–220. doi: 10.1109/IC3A48958.2020.233300.

[13] Apoorva Mysore Suresha, H. Behara, and G. More, "Sentiment Analysis on Amazon Product Reviews with Stacked Neural Networks," 2020, doi: 10.13140/RG.2.2.14746.67524.

[14] A. Rashid and C. Huang, "Sentiment Analysis on Consumer Reviews of Amazon Products," *International Journal of Computer Theory and Engineering*, vol. 13, no. 2, p. 7, 2021.

[15] G. V. A. Gutiérrez, "A Comparative Study of NLP and Machine Learning Techniques for Sentiment Analysis and Topic Modeling on Amazon," *IJCSENET*, vol. 9, no. 2, pp. 159–170, Apr. 2020, doi: 10.21817/ijcsenet/2020/v9i2/200902007

[16] B. Gezici, N. Bolucu, A. Tarhan, and B. Can, "Neural Sentiment Analysis of User Reviews to Predict User Ratings," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, Samsun, Turkey, Sep. 2019, pp. 629–634. doi: 10.1109/UBMK.2019.8907234.

[17] A. Dadhich, and B. Thankachan, "Sentiment Analysis of Amazon Product Reviews Using Hybrid Rule-based Approach," *IJEM*, vol. 11, no. 2, pp. 40–52, Apr. 2021, doi: 10.5815/ijem.2021.02.04.