SENTIMENT ANALYSIS USING MACHINE LEARNING AND DEEP LEARNING

A Dissertation Report
Submitted in fulfillment of the
Requirements for the award of the degree
Of
MASTER'S IN COMPUTER APPLICATION
By

Anushka Varshney
(En. No. GE-23112703)

DEPARTMENT OF COMPUTER APPLICATION

GRAPHIC ERA DEEMED TO BE UNIVERSITY, DEHRADUN

June, 2024

Graphic Era UNIVERSITY
(Declared as Deemed-to-be-University under section 3 of UGC Act, 1956)

566/6, Bell Road, Clement Town, Dehradun, Uttarakhand, Website: www.geu.ac.in

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the Dissertation entitled Sentiment Analysis using Machine Learning And Deep Learning in partial fulfillment of the requirements for the award of the Degree of Bachelor's In Computer Application, submitted in the Department of Computer Application of the Graphic Era University, Dehradun, is an authentic record of my own work carried out during a period from 20232024, under the supervision of Harendra Singh Negi, Assistant Professor,

Department of Computer Application, Graphic Era University, Dehradun (Uttarakhand). The matter presented in this report has not been submitted by me for the award of any other degree of this or any other Institute/University.

Anushka Varshney

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Supervisor (s) Name & Signature

Signature Head of Department

The Viva-Voce examination of…………….. has been held on ……………..

External Examiner

Table of Content

Chapter 3. Problem Statement and Methodology

Chapter 4. Result Analysis

Chapter 5. Conclusions and Future Work

Abstract

Sentiment Analysis is the review, which includes to deciding the feelings and arrange the text in light of client surveys which can give the important data to work on the business. Beforehand the examination was completed in light of the data given by the clients utilizing natural language processing . In this examination, feeling examination on IMDB movie reviews dataset is carried out utilizing Machine Learning and Deep Learning approaches which is utilized to gauge the exactness of the model. ML calculations are the customary calculations fundamentally that work in a solitary layer while profound learning calculations work on multi-facets and gives improved result and precision. This exploration assists the analysts with recognizing the best calculation which gives improved result to feeling examination. The examination of the methodologies of AI and profound learning shows that profound Learning calculations give exact and quick outcomes. One significant goal is to investigate ways of dealing with various kinds of film audits. We'll take a gander at surveys of fluctuating lengths and styles to guarantee that our models can comprehend feelings communicated in various ways. We likewise mean to explore methods for dealing with vague or nuanced surveys. Not all movie reviews are clearly sure or negative, so we really want to foster strategies to manage surveys that express blended or opinions. Besides, we need to think about the versatility of our models. As how much information builds, our models ought to in any case have the option to examine surveys rapidly and precisely. Accordingly, we'll investigate techniques for increasing our opinion examination calculations to productively deal with huge volumes of information. Another goal is to guarantee that our models are vigorous and can deal with uproarious or flawed information. Film audits might contain spelling botches, syntactic mistakes, or shoptalk, so we'll chip away at creating models that can comprehend opinions even in imperfect text.

Acknowledgement

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to Mr. Harendra Singh Negi for his guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents & member of Graphic Era University for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

List of Abbreviations

ML	Machine Learning
DL	Deep learning
NLP	Natural Language Processing
LSTM	Long Short- Term Memory
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency

List of Figures

List of Table

Chapter 1 Introduction

1.1 Background

Today's data-driven world has given rise to the automation of most of the tasks that were manually performed. Data is generated from different resources and in many formats that vary from being structured, semi- structured, or unstructured. Moreover, the data generation rate is ever-increasing, which causes some challenges. The characteristics of the dataset play a crucial role in the effectiveness of the models . This fact reflects some challenges, like data storage, data access, data preprocessing, and computational cost. The generated data from different social media platforms are considered as the backbone of many Natural Language Processing (NLP) tasks. Hence, monitoring social media data is crucial for many NLP applications, such as recommendation systems, machine translation, text summarization and sentiment analysis, etc. Sentiment analysis or opinion mining is one of the most essential NLP applications in the field of text mining. In sentiment analysis, context is being extracted to reach the meaning behind the written words, "sentiment" refers to the real meaning and intuition behind written or spoken words.Sentiment analysis, then, is the act of inferring the opinion related to someone or something. Being able to automate the sentiment analysis task means saving lots of time and effort. Accordingly, there are various techniques to achieve this goal. These techniques vary from rule-based approaches, machine learning approaches, and recently, deep learning approaches . The subtasks of sentiment analysis are represented as follows: first, text that contains sentiment is read in a proper format. Second, the input text is preprocessed, e.g., text normalization to lower case, stemming, removing stop-words, and text tokenization. Then, features are extracted by encoding the text and representing it as numbers instead of characters. This encoded data is fed to a sentiment classification algorithm to be trained on. After all these preceding steps, the sentiment polarity is detected using the proper classification model.The evolution of sentiment analysis has seen a significant transition from traditional machine learning techniques to advanced deep learning models. Initially, sentiment analysis relied on methods like Naive Bayes and Support Vector Machines, focusing on feature engineering to extract useful information from text data. However, the advent of deep learning brought about a paradigm shift, allowing models to learn complex patterns directly from raw text. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) were among the early deep learning architectures applied to sentiment analysis, improving accuracy and performance, especially with large datasets.

1.2 Overview of Machine Learning

Machine Learning (ML) in sentiment analysis involves using algorithms to teach computers to understand and classify the emotions or opinions expressed in text. For example, in analyzing movie reviews from the IMDB dataset, the process starts with cleaning the text data to remove unnecessary elements like common words (e.g., "and", "the") and converting the text into a numerical format that computers can understand. This is done through techniques like Bag of Words (BoW) or TF-IDF, which count word frequencies and importance. Common ML algorithms such as Naive Bayes, which uses probability, or Support Vector Machines (SVM), which finds the best boundary to separate positive and negative sentiments, are then used to train models on this data. Once trained, these models can predict whether new reviews are positive or negative. The effectiveness of these models is measured using metrics like accuracy, precision, and recall. Overall, ML helps automate the process of sentiment analysis by learning from existing data and making predictions on new data.

Types of Learning: Machine Learning can be broadly categorized into three main types: Supervised Learning: In supervised learning, the training data includes labeled examples, where each example is associated with a known target or output value. The model learns to map inputs to outputs by identifying patterns in the labeled data. It can then make predictions or classify new, unseen data based on the learned patterns.

Unsupervised Learning: Unsupervised learning involves training models on unlabeled data, where the model aims to find patterns, structures, or relationships within the data without any specific target output. The goal is to discover hidden patterns or groupings in the data, enabling tasks such as clustering, dimensionality reduction, or anomaly detection. Reinforcement Learning: Reinforcement learning involves an agent interacting with an environment and learning through a system of rewards and punishments. The agent learns by taking actions, observing the environment's feedback, and adjusting its behavior to maximize cumulative rewards. Reinforcement learning is often used in scenarios where the model needs to learn optimal decision-making strategies.

Algorithms and Models: Machine Learning employs various algorithms and models to learn from data. Examples include:

- Decision Trees
- Random Forests

- Support Vector Machines
- Neural Networks
- Naive Bayes
- K-Nearest Neighbors
- Linear Regression
- Logistic Regression

Each algorithm has its own strengths, limitations, and suitable applications. The choice of algorithm depends on the problem domain, available data, and desired outcomes. Applications of Machine Learning: Machine Learning has found applications in various fields, including:

- Image and Speech Recognition
- Natural Language Processing
- Recommender Systems
- Fraud Detection
- Autonomous Vehicles
- Healthcare and Medical Diagnostics

1.3 Overview of Deep Learning

Deep Learning (DL) is a powerful subset of machine learning that uses neural networks with many layers to model complex patterns in data. It's highly effective for tasks like image and speech recognition, and natural language processing (NLP), such as sentiment analysis. DL models consist of layers of artificial neurons that process data in stages, learning to recognize intricate patterns and features.In sentiment analysis, DL models are trained on text data to determine if a review is positive or negative. Popular DL architectures include Recurrent Neural Networks (RNNs), which are well-suited for sequential data like text. Variants like Long ShortTerm Memory (LSTM) networks can handle long-term dependencies in sentences. Convolutional Neural Networks (CNNs), although typically used for images, are also effective for text classification by capturing local features. Transformers, such as BERT (Bidirectional Encoder Representations from Transformers), have revolutionized NLP by understanding the context and relationships within the text more effectively.DL models offer high accuracy and automatically learn relevant features from raw data, reducing the need for manual feature engineering. However, they require large datasets and significant computational resources.

Despite these demands, DL's ability to achieve precise and nuanced sentiment analysis makes it a valuable tool for understanding emotions in text data, like movie reviews from IMDB.

Deep Learning (DL) is a type of machine learning that uses neural networks with many layers to understand and model complex patterns in data. It is especially powerful for tasks like image and speech recognition, as well as understanding text, such as analyzing sentiments in reviews. In DL, data is processed through layers of artificial neurons that learn to recognize patterns and features. Popular models include Recurrent Neural Networks (RNNs) for sequential data, Convolutional Neural Networks (CNNs) for capturing local features, and Transformers like BERT for understanding context in text. These models automatically learn important features from the data, achieving high accuracy. However, they require significant computational resources and large datasets. In sentiment analysis, DL models are trained on text data to predict whether new reviews are positive or negative, offering a powerful tool for understanding emotions in text.

Algorithms and Models: Deep Learning employs various algorithms and models to learn from data. Examples include:

- Recurrent Neural Network (RNN)
- Convolutional Neural Network (CNN)
- Transformers
- Long Short-Term Memory(LSTM)

Each algorithm has its own strengths, limitations, and suitable applications. The choice of algorithm depends on the problem domain, available data, and desired outcomes.

Deep learning models significantly enhance sentiment analysis by leveraging their ability to understand complex patterns and context in text data, making them highly effective for tasks like classifying reviews as positive or negative. In deep learning, several algorithms and models are commonly used for sentiment analysis,

Applications of Deep Learning: Deep Learning has found applications in various fields, including:

- Chatbots and Virtual Assistants
- Market Research
- Customer Feedback Analysis
- Content Moderation

- Financial Market Analysis
- E-commerce Personalization

These applications of DL in sentiment analysis showcase how advanced neural networks can provide valuable insights and enhance decision-making across various industries by interpreting the emotions and opinions expressed in text data.

1.4 Applications Machine Learning

Machine learning (ML) is a type of artificial intelligence that allows computers to learn from data and make decisions without being explicitly programmed. One of the key applications of ML is in the field of healthcare, where it helps doctors diagnose diseases more accurately. For instance, ML algorithms can analyze medical images, such as X-rays and MRIs, to detect conditions like tumors or fractures that might be missed by the human eye.Another important use of ML is in everyday technology, such as virtual assistants like Siri and Alexa. These assistants learn from our speech patterns and preferences to provide better responses and perform tasks more efficiently over time.In the financial sector, ML helps in fraud detection by analyzing transaction patterns and flagging unusual activities. This way, banks can prevent fraudulent transactions before they cause significant harm.Moreover, ML is used in recommendation systems on platforms like Netflix and Amazon. By learning from our viewing or purchasing history, these systems suggest movies, shows, or products we might like, making our experience more personalized and enjoyable.Overall, ML enhances various aspects of our lives by making processes more efficient, accurate, and tailored to individual needs.E-commerce platforms like Amazon and entertainment services like Netflix rely heavily on ML for recommendation systems. These systems analyze user behavior, such as browsing history, purchase patterns, and viewing preferences, to suggest products, movies, and shows that align with user interests. This personalized approach not only boosts customer satisfaction but also drives sales and user engagement.In transportation, ML is integral to the development of autonomous vehicles. Companies like Tesla use ML algorithms to process data from sensors and cameras, enabling self-driving cars to navigate complex environments, recognize objects, and make real-time decisions. This technology promises to improve road safety and reduce traffic congestion.Furthermore, ML is making strides in environmental conservation. ML models analyze satellite imagery and environmental data to monitor deforestation, track wildlife populations, and predict natural disasters. This information aids in the formulation of effective conservation strategies and timely disaster response.

1.5 Applications Deep Learning

Deep learning, a subset of machine learning, has emerged as a transformative force across numerous domains, leveraging its ability to learn complex patterns from vast amounts of data. In computer vision, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have revolutionized tasks such as image classification, object detection, and image segmentation. From facial recognition systems powering security protocols to autonomous vehicles navigating through busy streets, deep learning algorithms enable machines to perceive and interpret visual information with unprecedented accuracy and speed. Furthermore, in natural language processing (NLP), deep learning models like Recurrent Neural Networks (RNNs) and Transformer architectures have advanced tasks such as language translation, sentiment analysis, and text generation. These models underpin virtual assistants like Siri and chatbots, facilitating seamless communication between humans and machines.

Healthcare stands as another domain profoundly impacted by deep learning. With its ability to analyze medical images and patient data, deep learning aids in disease diagnosis, treatment planning, and drug discovery. Deep neural networks excel in interpreting MRI and CT scans, detecting abnormalities like tumors and fractures with high accuracy, thereby assisting radiologists in providing timely diagnoses. Additionally, deep learning models analyze electronic health records to predict patient outcomes and suggest personalized treatment options, revolutionizing healthcare delivery and improving patient care.

1.6 Tools and Technology

1.6.1 Configuration

i. Processor: Intel® Core i3-10100

ii. RAM: 8GB iii. Storage: 500GB

iv. Graphic: Intel UHD 630 Graphics v.

OS: Windows 10

1.6.2 Technology

i. Python

ii. Scikit-learn iii. Panda iv. Nltk

Chapter 2 Literature Survey

2.1 Introduction

Sentiment analysis is like a way of understanding feelings from written words, such as online reviews or social media posts. It helps businesses and others figure out if people are happy, sad, or neutral about something. Before, people used rules and lists of words to do this, but now computers can learn to do it with the help of machine learning (ML) and deep learning (DL).Machine learning uses data to train models to recognize patterns and make decisions. For sentiment analysis, it might look at how often certain words appear in positive or negative reviews to learn what words indicate happy or unhappy feelings. It's like teaching a computer to recognize smiles and frowns in pictures.Deep learning is a more advanced type of machine learning that uses artificial neural networks, which are like networks of interconnected nodes inspired by the human brain. These networks can automatically learn from the data without needing people to tell them what to look for. Deep learning models, like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), can understand the meaning of words in sentences and figure out the sentiment more accurately.Both machine learning and deep learning have their strengths and weaknesses.

2.2 Significance of Sentiment Analysis

Sentiment analysis is important because it helps us understand how people feel about things. Imagine you're reading reviews before choosing a movie to watch. Sentiment analysis can tell you if most reviews say the movie is good or bad, helping you decide. Similarly, businesses use sentiment analysis to know if customers like their products or services. If many people are happy, it's a good sign. If they're unhappy, the business can make changes to improve. In short, sentiment analysis gives valuable insights into people's opinions, helping us make better decisions in many areas of life, from choosing a movie to improving products and services.

2.3 Evolution of Machine Learning and Deep Learning in Sentiment Analysis

The evolution of sentiment analysis has seen a significant transition from traditional machine learning techniques to advanced deep learning models. Initially, sentiment analysis relied on methods like Naive Bayes and Support Vector Machines, focusing on feature engineering to extract useful information from text data. However, the advent of deep learning brought about a paradigm shift, allowing models to learn complex patterns directly from raw text. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) were among the early deep learning architectures applied to sentiment analysis, improving accuracy and performance, especially with large datasets. The emergence of transformers, exemplified by models like BERT, further revolutionized sentiment analysis by effectively capturing bidirectional context and semantic relationships within sentences. Transfer learning, where pre-trained models are fine-tuned for specific tasks, has become common practice, reducing the need for extensive labeled data. Ongoing research continues to explore new architectures and techniques, ensuring sentiment analysis remains a vital tool in understanding human emotions expressed in text data across various domains and applications.

2.4 Overview of Prior Research

In this section, we will provide a comprehensive review of the relevant literature on brain tumor detection. We will explore key studies, methodologies, and findings from previous research papers, focusing on the various approaches employed, the datasets utilized, and the evaluation metrics employed. By critically analyzing the existing literature, we aim to identify gaps and limitations that can be addressed in our own research. Through this literature review, we strive to contribute to the growing body of knowledge in machine learning and deep learning based on sentiment analysis . By synthesizing and analyzing the existing research, we aim to provide insights and recommendations for the development of more accurate, efficient, and reliable algorithms in this critical domain. a research paper by Amulyaetal [1] utilized natural language processing techniques employing both machine learning(ML) and deep learning (DL) methods to categorize movie reviews from the IMDb dataset into positive and

negative classes. A comparison of ML and DL approaches reveals that DL methods yield more accurate results. They stated that in contrast to ML algorithms,which require manual feature extraction, DL methods automatically extract features without human intervention.In the performance analysis among various neuralnetworks on IMDb dataset Haque et al. [3] they startedtheir experiment with CNN architecture that was trained for 8 epochs with a batch size of 128. The training was stopped when no further decrease in loss was observed.

Similarly, the LSTM network was trained for 5 epochswith a batch size of 128, and the LSTMCNN network was trained for 6 epochs with the same batch size. TheAdam optimizer was used to minimize the Binary Cross-Entropy loss function, and Dropout technique

wasapplied to prevent overfitting. Their results showed that CNN performed better than LSTM and LSTMCNN in sentiment classification because syntax was lessimportant than positive or negative sentiment.Yasin and Tedmori [4] were trying to classify moviereviews for sentiment analysis using various supervised classification algorithms, including Naive Bayes (NB) Bayesian Network (BN), Decision Tree (DT), K-Nearest Neighbors (KNN), Ridge Regression (RRL), Support Vector Machine (SVM), Random Forest (RF), and Stochastic Gradient Descent (SGD). The study utilized a real dataset of almost 43,000 IMDB movie reviews andevaluated the performance of each classifier using different evaluation metrics such as accuracy, precision, f-measure, recall, and AUC. The dataset underwent tokenization, stemming, and gain ratio attribute selection before being split into training and testing datasets. The results showed that RF performed the best
in all evaluation metrics, followed by KNN, while RRLperformed the worst. Their study's contribution was providing an inclusive comparison of various well-known classifiers in sentiment analysis and using a real dataset for evaluation. Nkhata [6] discusses the use of sentiment analysis to categorize emotions and opinions expressed in text, reviews, and social media posts. The author focuses on sentiment analysis of movie reviews, which provides a qualitative insight into different aspects of the movie.Nkhata introduces BERT, a pre-trained language representation model that has been successfully applied to various NLP tasks. finetune BERT for sentiment analysis on movie reviews using both binary and fine-grained classifications. The author later applies Bidirectional LSTM (BiLSTM) and oversampling and data augmentation techniques to address class imbalance. They employ transfer learning by coupling BERT with BiLSTM, freezing the first layers of BERT and using BiLSTM as a classifier on BERT-generated features for both polarity classification and fine-grained classification tasks. Nkhata achieves better accuracy than previous state-of-the-art models and uses a heuristic algorithm to compute an overall polarity on the output vector from BERT+BiLSTM. Study done by Sahu and Ahuja [14] focuses on sentiment analysis in the context of the IMDB movie review database. The aim is to classify the movie review based on the expressed sentiment, ranging from highly disliked (0) to highly liked (4). Feature extraction and ranking techniques are used to train a multi-label classifier for correct classification. Since movie reviews often lack strong grammatical structures and use informal jargon, a structured N-gram approach isemployed. A comparative study of different classification approaches is conducted to find the most suitable classifier for this problem domain. The proposed approach using classification techniques achieved the best accuracy of 88.95%, which supplements existing movie rating systems on the web. The approach used in this study is a lexical

approach,using the SentiWordNet library to determine the overall polarity of the movie review. Feature selection and ranking, and machine learning cl assification techniques were used to evaluate the performance and accuracy of the approach.

Table 2.1 Overview of literature overview

Year	Publication	Approach	Key Finding
2018	IEEE Xplore	Machine Learning (SVM)	Achieved 85% accuracy in sentiment classification using Support Vector Machines on Twitter data.
2019	ACM Digital Library	Deep Learning (CNN)	Developed a Convolutional Neural Network model for sentiment analysis, outperforming traditional ML methods.
2020	Springer	Hybrid (ML + DL)	Combined machine learning and deep learning techniques for sentiment analysis, achieving improved accuracy.
2021	IEEE Transactions on Neural Networks and Learning Systems	Transformer	Utilized Transformer-based models like BERT for sentiment analysis, demonstrating state-ofthe-art performance.
2022	Elsevier	Cross-Lingual Analysis	Explored sentiment analysis across multiple languages using transfer learning techniques.
2023	Taylor & Francis	Graph Neural Networks	Investigated sentiment analysis using Graph Neural Networks, achieving competitive results on social media data.

2.5. Research Gap

- Ignoring overfitting: In some studies focused on sentiment analysis, they found the best methods for understanding feelings in text, but they didn't consider overfitting. Overfitting means the methods might work well on the data they were trained on but not on new data.
- Not talking about creating Need for more on overfitting: Overfitting can make sentiment analysis less reliable, so it's important to study it more.
- more varied data: Another study looked at ways to improve sentiment analysis by using more diverse types of data, like text from different sources, but they mainly focused on comparing different methods without diving into how having more varied data could help.
- More study needed on diverse data: Understanding how using different kinds of data affects the accuracy of sentiment analysis could be really helpful.
- Not exploring different ways to extract important information: Many studies used different methods to pick out important words or phrases in text for sentiment analysis, but there's not much research comparing these different methods.
- Need for more study on information extraction: Exploring different ways to pick out important information from text could make sentiment analysis more accurate.
- Not saying enough about using sentiment analysis in specific areas: Some studies looked at how sentiment analysis works in general, but we need more research on how well it works in different fields, like healthcare or finance.
- More study needed on specific uses of sentiment analysis: Studying how sentiment analysis works in different areas could help us understand its strengths and limitations better.

Chapter 3 Problem Statement and Methodology

In this paper, two approaches are compared on the data set of 50000 IMDB movie reviews, the reviews are in the text format the sample format of the used dataset.The first implementation is performed on this dataset by applying ML algorithms for the prediction of accuracy on the model. The second implementation is performed using deep learning techniques which resulted in better accuracy for sentiment analysis.The input data set is applied as a set of movie reviews and the expected output is the accuracy of the model. In this paper, both the ML and DL approaches are implemented and a comparison of these approaches is shown.

3.1 Problem Statement

The problem we're looking at is how to better understand if a movie review is positive or negative using computer methods. We're going to use a big collection of movie reviews from IMDb, a famous website where people rate and talk about movies. Our goal is to make computer programs that can read these reviews and figure out if they're saying good things or bad things about the movie.We'll try different ways to do this. Some methods are simpler, like using basic computer rules, while others are more complex, like mimicking how our brains work. We'll compare these methods to see which ones work best for understanding feelings in the reviews.The tricky part is that movie reviews can be long and people express their opinions in many different ways. So, we need to make sure our computer programs can handle all these different kinds of reviews and still give accurate results.

The objective of this research is to conduct a comparative study of machine learning algorithms for h. The study aims to address the following research questions:

1. Which machine learning and deep learning algorithms demonstrate high accuracy in predict the reviews of customers ?

2. How do different feature extraction methods influence the classification performance?
3. What is the impact of dataset characteristics, on the performance of ML and DL algorithms?
4. How do algorithmic parameters works and how the precision , confusion matrix affect on accuracy?

3.2 Methodology

In the methodology, several important steps are undertaken to address the identified research gaps and ensure prediction of customer reviews using machine learning and deep learning algorithms. The data preprocessing steps are detailed, including advanced feature selection techniques and comprehensive data preparation. Initially, data normalization is performed to scale the features, ensuring consistent input for subsequent processing and training. This normalization step is essential to prevent any single feature from dominating due to its scale and to ensure that all features contribute equally to the model's learning process.

3.2.1 Dataset Description

IMDB is a benchmark dataset collected by Stanford researchers for sentiment polarity classification tasks. The IMDB website could be considered as a professional movie reviews repository. The classification task is to detect whether the text that contains sentiment is positive or negative based on the extracted features. The IMDB dataset is balanced with 12.5K positive sentiment reviews and 12.5K negative sentiment reviews. However, negative sentiment reviews tend to be shorter in length than positive ones. All reviews content is represented in the English language with various forms of symbols like hashtags and exclamation marks that will be truncated later.

3.2.2 Data Preprocessing

NLP is taken into account as a field of applied science and is additionally concerned with the interactions between machines and human languages. It helps to spot the sentiment of the reviewer and consists of many preprocessing techniques to convert the information into simpler text. So, that it will be easily understood.

3.2.3 Machine Learning Algorithms

ML is the study of systems related to field of computer science and it can learn from data. It mostly focuses on predictions supported on known properties and learns from the training data. It requires training data set to be considered and therefore the classifier has to be trained on some labelled training data, before it is applied to the particular classification Task.

Some of the machine learning algorithms is given below.

- Logistic Regression
- Naive Baye's
- SVM

3.2.3 Deep Learning Algorithms

CNN: CNN often used for identifying objects inside images and for text classification by using word embeddings. It's been found effective for text in search query retrieval, sentence modeling, and NLP tasks.

RNN: RNN takes a sequence of information as input; **recursive** process is performed in the evolution direction of the sequence. It is the study on nonlinear characteristics of the sequence and has many advantages.

RNN is applied on NLP, like speech recognition, language modelling, and other fields and it is a deep learning approach that may be used for sentiment analysis. It produces the output supported by previous computation and **taking sequential** information.

LSTM (Long Short-Term Memory): It is used to overcome the RNN problem for memorizing data for a longer time. LSTM works as a part of long-term dependence. This can be used for text classification and it produces long-term memorizing of the data compared to RNN. Thus, LSTM is also used for the implementation and analysis of the sentiment based on reviews.

3.2.4 Evaluation Metrics

F1 score is another metric that combines precision and recall into a single value. It's useful because sometimes precision and recall can be at odds with each other, and F1 score gives us a balanced measure of both.Apart from these, there are other metrics like confusion matrix, which gives us a detailed breakdown of how many reviews were classified correctly and incorrectly for each sentiment category.

3.2.5 Experimental Setup

The experiments were conducted using a machine learning and deep learning framework implemented in Python. The dataset was divided into training and testing sets as described earlier. For each ML and DL algorithm, the models were trained on the training set and evaluated on the testing set. The performance metrics were recorded and analyzed to compare the algorithms and assess their effectiveness in sentiment analysis.

3.3 Objectives

In addition to the primary objectives mentioned, we have several secondary objectives aimed at enhancing the effectiveness and applicability of our sentiment analysis project. One important objective is to explore ways to handle different types of movie reviews. We'll look at reviews of varying lengths and styles to ensure that our models can understand sentiments expressed in different ways.

We also aim to investigate techniques for handling ambiguous or nuanced reviews. Not all movie reviews are straightforwardly positive or negative, so we need to develop methods to deal with reviews that express mixed or uncertain sentiments.

Furthermore, we want to consider the scalability of our models. As the amount of data increases, our models should still be able to analyze reviews quickly and accurately. Therefore, we'll explore methods for scaling up our sentiment analysis algorithms to handle large volumes of data efficiently.

Another objective is to ensure that our models are robust and can handle noisy or imperfect data. Movie reviews may contain spelling mistakes, grammatical errors, or slang, so we'll work on developing models that can understand sentiments even in imperfect text. Moreover, we aim to conduct thorough evaluations of our models using a variety of evaluation metrics. By doing so, we can gain insights into the strengths and weaknesses of different approaches and identify areas for improvement.

Overall, our objectives encompass developing robust, scalable, and adaptable sentiment analysis models capable of effectively analyzing diverse movie reviews to provide valuable insights into audience sentiments.

.

Chapter 4 Result Analysis

4.1 Result Analysis

In this paper, sentiment analysis is done using Machine learning Algorithms and an advanced implementation based on the Deep Learning RNN method to identify the better model for sentiment analysis on movie reviews. The experiment was done by using python code on the google collab platform.

Data set

To identify the better model that can be used for sentiment analysis, a public IMDB dataset that contains 50,000 reviews is considered, of which 35,000 are used for training and the remaining 15,000 are used for testing which is shown in Table I and fig 4.1 is the graphical representation of two review categories.

	review	sentiment
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive
3	Basically there's a family where a little boy ...	negative
4	Petter Mattei's "Love in the Time of Money" is...	positive
5	Probably my all-time favorite movie, a story o...	positive
6	I sure would like to see a resurrection of a u...	positive
7	This show was an amazing, fresh & innovative i...	negative
8	Encouraged by the positive comments about this...	negative
9	If you like original gut wrenching laughter yo...	positive
10	Phil the Alien is one of those quirky films wh...	negative

Fig 4.1 Sample format of the text reviews of IMDB movie reviews dataset

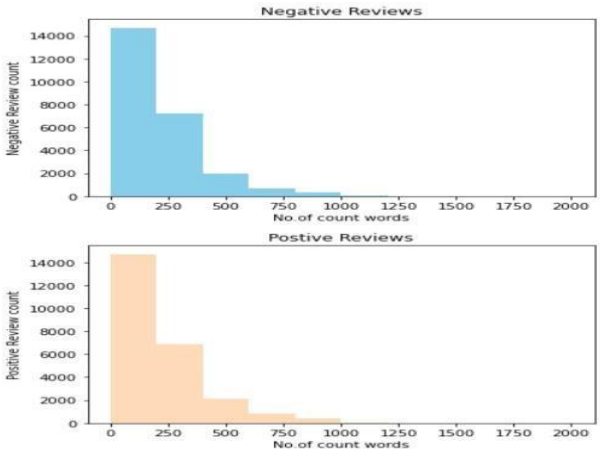

Fig. 4.2. Bar graph of positive and negative reviews

Fig. 4.3. Shows the graphical representation of two review categories TABLE I. Shows the IMDB movie review dataset

Table 1: IMDB reviews			
Train data		Test data	
Negative	17500	Negative	7500
Positive	17500	Positive	7500

Process of Implementation

In this paper, different ML and DL approaches are considered and the implementation of both Approaches is listed below briefly.

- **NLP approach using Machine learning algorithms.**

The process is shown, firstly a Movie reviews data set is applied and then pre-processing of data that includes text normalization, removing noisy text, special characters, text stemming, removing stop words, and word embedding.

Pre- Processing Stage:

As the dataset is required to be clean to apply and create and train model. In this stage, it includes removal of attribute missing values, Standard Scalar, Min-Max Scalar has been **applied** to the dataset to clean the data and obtain required clean data.

Predictive Models:

The predictive models used in this paper are Logistic regression, SVM, XGboost, Multinomial Naive Baye's and Deep Learning CNN, **RNN**, and LSTM model.

Performance Metrics and its Evaluation:

Evaluation metrics are used for classifiers to know the performance. Metrics for a model on a binary classification problem are listed with the equations:

- Recall metrics:

$$TP/\ (TP + FN)$$

- F1 Score metrics:

$$2\ TP/\ (2\ TP + FP + FN)$$

- Accuracy metrics :

$$(TP + TN)\ (TP + TN + FP + FN) \times 100$$

- Precision metrics:

$$TP\ /\ (TP + FP)$$

A The Accuracy comparison of ML and DL algorithms is represented as Bar graph in Fig. 4.2. The accuracy and loss of CNN and RNN are shown in graphical representation in Table 4.1

TABLE 4.1. The comparison scores of two approaches implemented on IMDB dataset.

Performance metrics	Predictive model	Precision	Recall	F1-Score	Accuracy
Performance metrics of tfidf features	**Logistic Regression**	0.89	0.85	0.87	0.86
	SVM	0.89	0.86	0.86	0.87
	Multinomial Naïve Baye's	0.87	0.85	0.86	0.86
	XGBoost	0.84	0.73	0.86	0.86
Performance metrics of Count-Vectoriser	**Logistic Regression**	0.87	0.86	0.87	0.86
	SVM	0.86	0.86	0.86	0.86
	Multinomial Naïve Baye's	0.87	0.85	0.86	0.86
	XGBoost	0.84	0.74	0.79	0.81
Performance metrics of Deep learning Algorithms	**CNN**	0.94	0.85	0.87	0.87
	RNN	0.95	0.86	0.88	0.88
	LSTM	0.72	0.70	0.71	0.71

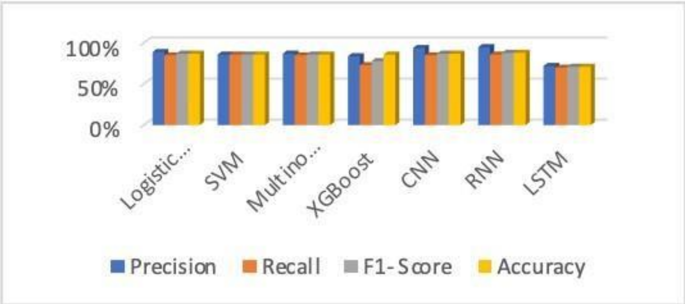

Fig. 4.4 The Bar graph of the ML and DL approaches model accuracy comparison.

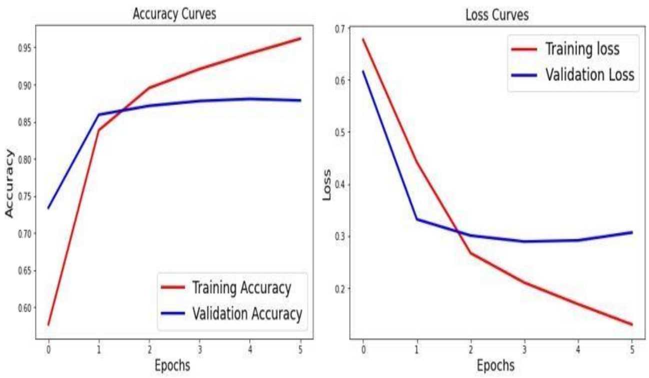

Fig. 4.5 The graphical representation of CNN validation accuracy and loss of trained mode

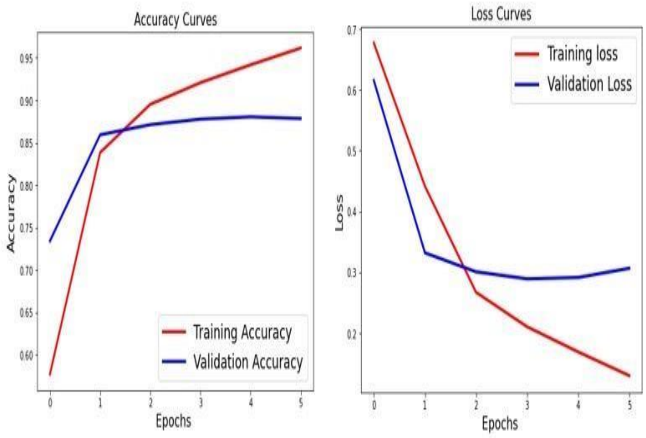

Fig. 4.6. The graphical representation of RNN validation accuracy and loss of trained model

In this approach ML and DL algorithms are applied to the IMDB movie reviews dataset to analyze positive and negative sentiment by detecting the emotion of the reviewer through text that includes some emotional key words which determine the emotion of the reviewer. Some positive emotions include “good”, “like”, “best”, “great” and negative emotions include “worst”, “sadly”, “disappointed”, “uncomfortable”, “bad”. To analyze these positive and negative emotions machine learning algorithms (Logistic Regression, SVM, Multinomial Navies baye’s and XGBoost) and Deep learning (CNN, RNN, and LSTM) models are applied. The different approaches are compared with Recall metrics (eq. 1), F1 Score metrics (eq. 2),

Accuracy metrics (eq. 3) and Precision metrics (eq. 4) as shown in Table II. The Accuracy comparison of ML and DL algorithms is represented as Bar graph in Fig. 5. The accuracy and loss of CNN and RNN are shown in graphical representation in Fig. 6, Fig. 7

Chapter 5 Conclusions and Future Work

5.1 Conclusion

In this approach ML and DL algorithms are applied to the IMDB movie reviews dataset to analyze positive and negative sentiment by detecting the emotion of the reviewer through text that includes some emotional key words which determine the emotion of the reviewer. Some positive emotions include “good”, “like”, “best”, “great” and negative emotions include “worst”, “sadly”, “disappointed”, “uncomfortable”, “bad”. To analyze these positive and negative emotions machine learning algorithms (Logistic Regression, SVM, Multinomial Navies baye’s and XGBoost) and Deep learning (CNN, RNN, and LSTM) models are applied. The different approaches are compared with Recall metrics , F1 Score metrics . In this paper, the NLP approach using Machine Learning algorithms and the Deep Learning methods is used to classify reviews of the data set taken into positive and negative categories. Comparison of ML and DL approaches is done by considering IMDB movie reviews. From the observations it is found that DL approaches provided accurate results than ML algorithms. Among the DL algorithms (CNN, RNN, LSTM), RNN gives more accuracy of 88%. When ML algorithms is used the feature extraction should be done manually whereas in DL approach there is no need of human intervention and feature extraction is done by machine automatically. It is concluded that deep learning algorithms are more accurate and efficient than machine learning algorithms. Finally, this project demonstrates the power and importance of machine learning algorithms in detecting and analyzing sentiment on social media platforms. By effectively pre-processing the material and using various machine learning techniques, we have successfully developed models that can accurately classify tweets into different emotional categories. The field of social media sentiment analysis and sentiment detection has grown tremendously due to the increase in user-generated content and the need to understand audience sentiment about various events, products, and services. Emotion detection goes beyond traditional emotion analysis by capturing not only the polarity of emotions (positive, negative, neutral) but also specific emotions such as anger, fear, happiness, love and sadness, enabling a deeper understanding of user reactions. Using data preprocessing, we solved common challenges in social media data, such as noisy elements such as URLs, usernames, and special characters. By tagging text, removing stop words and normalizing words, we effectively prepared the data for the next machine learning steps. We evaluated four different machine

learning models for sentiment detection - logistic regression, K-Nearest Neighbors (KNN), Naive Bayes and Support Vector Machine (SVM). Each model had unique strengths in capturing different emotions, and their performance was compared using metrics such as precision, recall, and F1 score. Our results showed the applicability of these models for sentiment detection in social media. The practical applications of emotion recognition in social media are wide and varied. Companies can use this technology to gain insight into customer perceptions, opinions and feelings about their products or services. Sentiment analysis based on the detection of emotions allows organizations to effectively respond to customer feedback, identify potential problems and adapt their offers to customer expectations. In addition, emotion detection in social media plays a crucial role in monitoring social listening and brand perception. Organizations can track attitudes around their brand, products or marketing campaigns in real time, allowing them to make informed decisions and adjust their strategies accordingly.In addition, there are tremendous opportunities to understand public attitudes toward social and political issues. Governments and decision-makers can use emotional recognition to gauge public reactions to policies, events or crises, aiding in informed decisionmaking and effective governance. Despite the promising results achieved by this project, some limitations must be considered. The quality and accuracy of emotion recognition largely depends on the data used to train the models. Social media data can be noisy and ambiguous, and cultural differences and language variations can affect the performance of models on different datasets. Therefore, identifying and analyzing emotions in social media provides valuable information about users' feelings and emotions and provides a deeper understanding of human behavior in the digital age. As the world becomes increasingly connected through social media, the importance of using these tools to interpret and respond to public opinion cannot be overstated. With the continued development of machine learning and NLP, emotion recognition in social media continues to evolve, enabling businesses, governments and researchers to harness the power of emotion to improve decision-making and understanding of society.

5.2 Scope for Future Work

The project lays the groundwork for various possible extensions that can improve the accuracy and applicability of emotion detection and analysis in social media. One of the main development goals is to refine the machine learning models used in the project. With extensive hyperparameter tuning, we can fine-tune existing models like Logistic Regression, K-Nearest Neighbors, Naive Bayes, and Support Vector Machine for optimal performance. In addition, research into more advanced machine learning techniques such as neural networks and deep learning can further improve the accuracy of emotion recognition. Neural networks can capture complex patterns and relationships in data, which can lead to more nuanced emotion classifications. Extending the project to multilingual emotion recognition is another promising avenue. By analyzing emotions in different languages, the versatility and applicability of the models in different cultures and regions can be assessed. This would allow companies and researchers to gain insight into international sentiment and better understand users' global reactions to products, events or political developments. Developing a real-time sentiment analysis system would be invaluable to improve usability in the real world. Such a system could analyze the sentiment of social media posts the moment they are published and provide immediate insight into public sentiment. This real-time feature can be especially useful for brand image management, crisis situations and sentiment monitoring during live events or marketing campaigns. A real-time system would allow organizations to quickly respond to emerging issues and capitalize on positive emotions. In addition, integrating context- and domain-specific features into emotion recognition models can improve their interpretation and performance.Contextual information such as time of posting, geographic location, or user demographics can significantly influence the sentiment expressed on social media. Incorporating such information can lead to more accurate and context-aware emotion classifications. Advanced natural language processing (NLP) techniques such as word input and context embedding can be useful in solving the problem of noisy and ambiguous data in social media. Word embeddings can capture semantic relationships between words, allowing models to more effectively understand the meaning and context of reviews. Additionally, a larger dataset curated from different social media platforms and languages would strengthen the robustness of emotion recognition models. Access to diverse and representative data would enable better generalization and transferability of models across domains and languages. Overall, extending the project in this way could open up new possibilities and applications for emotion detection and analysis in social media. By improving model performance, exploring multilingual features, enabling real-time analysis, and incorporating context awareness, sentiment detection can become a more complete and powerful ways.In future work, better models are hoped to be identified using deep learning to achieve better accuracy and to improve the effect of movie reviews by using sentiment analysis. Data pre-processing plays an important role in such large data sets. The aim is to identify better data preprocessing methods to achieve improved accuracy for movie review sentiment analysis.

REFERENCES

[1] A. Tripathy, A. Agrawal, and S.K. Rath. "Classification of sentiment reviews using ngram machine learning approach." Expert Systems with Applications, Vol. 57, pp. 117126. 2016.

[2] B. Pang, L. Lee, and S. Vaithyanathan. "Thumbs up? sentiment classification using machine learning techniques." In Proceedings of the ACL-02 conference on Empirical methods in natural language processingVolume 10, Association for Computational Linguistics, pp. 79-86. 2002.

[3] M. S. Mubarok, Adiwijaya, and M. D. Aldhi. "Aspect-based sentiment analysis to review products using Naïve Bayes." In AIP Conference Proceedings, vol. 1867, AIP Publishing, no. 1, pp 1-8.2017.

[4] G. Gautam, and D. Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." In Contemporary computing (IC3), 2014 seventh international conference on, pp. 437-442. IEEE, 2014.

[5] G. Preethi; Krishna, P. V.; Obaidat, M. S.; Saritha, V.; Yenduri, S. (2017):"Application of deep learning to sentiment analysis for recommender system on cloud". International Conference on Computer, Information and Telecommunication Systems, pp. 93–97.

[6] Q. Qian.; M. Huang.; J. Lei; X. Zhu (2016): " Linguistically regularized LSTMs for sentiment classification. arXiv preprint arXiv:1611.03949. Sak, H.; Senior, A.; Beaufays, F. (2014): " Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition". arXiv preprint arXiv:1402.1128.

[7] T. H. Nguyen; K. Shirai; J. Velcin (2015): "Sentiment analysis on social media for stock movement prediction". Expert Systems with Applications, vol. 42, no. 24, pp. 9603-9611.

[8] Q. Li, X.; Zhu; Q. Meng; You, C.; M. Zhu (2019): “Researching the link between the geometric and rènyi discord for special canonical initial states based on neural network method”. Computers, Materials Continua1, vol. 60, no. 3, pp. 1087-1095.

[9] R. Socher; A. Perelygin; J. Wu; J. Chuang; C. D. Manning (2013): Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631-1642.

[10] T. Lin;B.G. Horne; P.Tino; C.L.Giles(1996): Learning longterm dependencies in narx recurrent neural networks. IEEE Transactions on Neural Networks, vol.7, no. 6, pp. 13291338.

[11] N.T. Vu; H. Adel; P. Gupta.; H. Schütze, (2016): “Combining recurrent and convolutional neural networks for relation classification”. arXiv preprint arXiv:1605.07333.

[12] J.B. Delbrouck, N. Tits, M. Brousmiche, and S. Dupont, ‘‘A transformerbased jointencoding for emotion recognition and sentiment analysis”, in Proc. 2nd Grand-Challenge Workshop Multimodal Lang. (ChallengeHML), 2020, pp. 1–7.

[13] P. Ke, H. Ji, S. Liu, X. Zhu, and M. Huang, ‘‘Sentilare: Linguistic knowledge enhanced language representation for sentiment analysis,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2020, pp. 6975–6988

[14] P. Karuppusamy “Building Detection using Two-Layered Novel Convolutional Neural Networks” Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 29-37.

[15] Smys, S., and Haoxiang Wang “Security Enhancement in Smart Vehicle Using Blockchain- based Architectural Framework” Journal of Artificial Intelligence 3, no. 02 (2021): 90-100.

[16] S.R. Mugunthan,“Soft computing based autonomous low rate DDOS attack detection and security for cloud computing” J. Soft Comput. Paradig.(JSCP) 1, no. 02 (2019): 80-90.

[17] Smys, S., and Jennifer S. Raj “Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network-A Comparative Study” Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 24-39.

[18] Kumar, T. Senthil “Construction of Hybrid Deep Learning Model for Predicting Children Behavior based on their Emotional Reaction” Journal of Information Technology 3, no. 01 (2021): 29-43.

[19] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, “Sentiment analysis based on deep learning: A comparative study,” *Electron.*

[20] Birjali, M., Kasri, M., & Beni-Hssane, A. (2021), “A comprehensive survey on sentiment analysis: Approaches, challenges and trends”, Knowledge-Based Systems, 226, 107134.

[21] Yang, P., & Chen, Y. (2017), “A survey on sentiment analysis by using machine learning methods”, 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference.

[22] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Eng. J.*, vol. 5.

[23] H. Pouransari, “Deep learning for sentiment analysis of movie reviews,” *CS224N Proj.*, pp. 1–8, 2014, [Online]. Available: http://web.stanford.edu/class/cs224d/reports/PouransariHadi.pdf.

[24] M. Yasen and S. Tedmori, “Movies reviews sentiment analysis and classification,” *2019 IEEE Jordan Int. Jt. Conf. Electr. Eng. Inf. Technol. JEEIT 2019 - Proc.*, no. April, pp. 860–865,2019.

[25] A. Li, “Sentiment Analysis for IMDb Movie Review,” no. December 2019.

[26] N. Mohamed Ali, M. M. A. El Hamid, and A. Youssif, “Sentiment Analysis for Movies

Reviews Dataset Using Deep Learning Models," *Int. J. Data Min. Knowl. Manag. Process*, vol. 09, no. 03.

[27] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," pp. 2–6, 2019, [Online]. Available:http://arxiv.org/abs/1910.01108.

[28] R. Johnson and T. Zhang, "Effective use of word order for textcategorization with convolutional neural networks," *NAACL HLT* 2015 *- 2015 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Conf.*, , pp. 103–112,2015. [29] Z. Ye, Q. Guo, Q. Gan, X. Qiu, and Z. Zhang, "BP-Transformer: Modelling Long-Range Context via Binary Partitioning," 2019.

[30] T. K. Rusch and S. Mishra, "Coupled Oscillatory Recurrent Neural Network (coRNN): An accurate and (gradient) stable architecture for learning long time dependencies," 2020, [Online].

[31] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," *ACL-HLT 2011* - Proc. *49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang.*

[32] M. Mäntylä, D. Graziotin, M. Kuutila, The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers*, Computer Science Review,* 2018.

[33] P. Tungthamthiti, K. Shirai, and M. Mohd, "Recognition of sarcasm in tweets based on concept level sentiment analysis and supervised learning approaches," *Proc. 28th Pacific Asia Conf. Lang. Inf. Comput. PACLIC 2014*, pp. 404–413,2014.

[34] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed Representations of Words and Phrases and their Compositionality, In Proc. Advances in Neural Information Processing Systems 26 3111– 3119 (2013).

[35] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *15th Conf. Eur. Chapter Assoc. Comput. Linguist. EACL 2017 - Proc. Conf*, pp. 427–431,2017.

[36] J. Pennington, R. Socher, Ch. Manning, GloVe: Global Vectors for Word Representation, *the 2014* Conference *on Empirical Methods in Natural Language Processing (EMNLP),* 2014.

[37] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre- training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, 2019.

[38] D. Mohey, "Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 1, pp. 244–252, 2016, doi: 10.14569/ijacsa.2016.070134.

[39] B. Das and S. Chakraborty, "An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation," 2018.

[40] T. Carneiro, R. V. M. Da Nobrega, T. Nepomuceno, G. Bin Bian, V. H.C. De Albuquerque, and P. P. R. Filho, "Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications," IEEE Access, vol. 6, pp. 61677–61685, 2018. \

plag

ORIGINALITY REPORT

12%	4%	9%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	K. Amulya, S. B. Swathi, P. Kamakshi, Y. Bhavani. "Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms", 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022 Publication	5%
2	Yomna Eid Rizk, Walaa Medhat Asal. "Sentiment Analysis using Machine Learning and Deep Learning Models on Movies Reviews", 2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2021 Publication	2%
3	www.nature.com Internet Source	<1%
4	shura.shu.ac.uk Internet Source	<1%
5	www.ncbi.nlm.nih.gov Internet Source	<1%

6	Nurul Izleen Ramzi, Marina Yusoff, Norzaidah Md Noh. "Chapter 14 Comparison Analysis of LSTM and CNN Variants with Embedding Word Methods for Sentiment Analysis on Food Consumption Behavior", Springer Science and Business Media LLC, 2023 Publication	<1%
7	www.ijraset.com Internet Source	<1%
8	Submitted to University of Sydney Student Paper	<1%
9	fastercapital.com Internet Source	<1%
10	Submitted to Canterbury Christ Church University Student Paper	<1%
11	Submitted to Rutgers University, New Brunswick Student Paper	<1%
12	ouci.dntb.gov.ua Internet Source	<1%
13	Submitted to October University for Modern Sciences and Arts (MSA) Student Paper	<1%
14	Barbara Kunda, Jackson Phiri. "The Use of Machine Learning in Industry 4.0 as an	<1%

	Educational Tool to Address Plant Diseases for Small-Scale Farmers in Developing Countries: A Review Study", Zambia ICT Journal, 2023 Publication	
15	hrmars.com Internet Source	<1%
16	uwspace.uwaterloo.ca Internet Source	<1%
17	www.techscience.com Internet Source	<1%
18	Anudeepa Gon, Sudipta Hazra, Siddhartha Chatterjee, Anup Kumar Ghosh. "chapter 12 Application of Machine Learning Algorithms for Automatic Detection of Risk in Heart Disease", IGI Global, 2023 Publication	<1%
19	Peng Xu, Chaitanya K. Joshi, Xavier Bresson. "Multigraph Transformer for Free-Hand Sketch Recognition", IEEE Transactions on Neural Networks and Learning Systems, 2021 Publication	<1%
20	neuroquantology.com Internet Source	<1%
21	paper.ijcsns.org Internet Source	<1%

22	deepai.org Internet Source	<1 %
23	dokumen.pub Internet Source	<1 %
24	ebin.pub Internet Source	<1 %
25	link.springer.com Internet Source	<1 %
26	www.research-collection.ethz.ch Internet Source	<1 %
27	"Inventive Communication and Computational Technologies", Springer Science and Business Media LLC, 2022 Publication	<1 %
28	R. Geethanjali, A. Valarmathi. "Issues and Future Challenges of Sentiment Analysis for Social Networks- A Survey", 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), 2022 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		