SENTIMENT ANALYSIS OF CUSTOMER FEEDBACK USING NLP TECHNIQUES

Syeda Sadika Anjum

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Syeda Sadika Anjum
Date of Birth	:	25 November 2002
Title	:	Sentiment Analysis of Customer Feedback Using NLP Techniques
Academic Session	:	2022-2025

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

Sadika

(Student's Signature)

(Supervisor's Signature)

221-35-1066
Student ID
Date:28/12/2025

Khalid Been Md. Badruzzaman Biplob
Name of Supervisor
Date: 28/12/2025

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

Department of Software Engineering

Faculty of Science and Information Technology

Supervisor Approval Form

Fall 2025	B.Sc. In SWE	Campus: DSC

Student Name	Student ID
Syeda Sadika Anjum	221-35-1066

Project/Thesis Information	
Project/Thesis Title	SENTIMENT ANALYSIS OF CUSTOMER FEEDBACK USING NLP TECHNIQUES
Type of work	Thesis

Supervisor information	
Supervisor Name	Mr. Khalid Been Badruzzaman Biplob
Supervisor Initial	KBB
Completed Credit till now	135
How many credits in this semester	12
Supervisor Consent	☑ Yes ☐ No

Supervisor Signature

APPROVAL

This thesis titled on "**Sentiment Analysis of Customer Feedback Using NLP Techniques**", submitted by **Syeda Sadika Anjum (ID: 221-35-1066)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- **Chairman**

Dr. S M Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 1**

A.H.M Shahariar Parvez
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 2**

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 3**

Khalid Been md. Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **External Examiner**

Dr. Md Sazzadur Rahman
Professor
Institute of Information technology
Jahangirnagar University, Bangladesh

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

____

(Supervisor's Signature)

Full Name :Khalid Been Md. Badruzzaman Biplob

Position :Senior Lecturer

Date :28/12/2025

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

Sadika

(Student's Signature)

Full Name : Syeda Sadika Anjum

ID Number : 221-35-1066

Date :28/12/2025

SENTIMENT ANALYSIS OF CUSTOMER FEEDBACK USING NLP TECHNIQUES

Syeda Sadika Anjum

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

December 2025

ACKNOWLEDGEMENT

We begin by extending our heartfelt gratitude to the Almighty for blessing us and enabling us to successfully complete our final year project/internship.

Our sincere appreciation goes to **Khalid Been Md. Badruzzaman Biplob, Senior Lecturer** of the Department of Software Engineering at Daffodil International University, Dhaka. Their extensive knowledge and keen interest in the field of "NLP" were instrumental in guiding us throughout this project. Their unwavering patience, scholarly guidance, continuous encouragement, dedicated supervision, constructive criticism, valuable advice, and thorough review of multiple drafts at every stage played a crucial role in the completion of this project.

We would also like to express our deepest thanks to , the Head of the Department of SWE, for his invaluable assistance in bringing our project to fruition. Our gratitude extends to all the faculty members and staff of the SWE, department at Daffodil International University.

We are grateful to our fellow classmates at Daffodil International University, who engaged in discussions and provided support during the course of our project.

Lastly, we would like to acknowledge and show our utmost respect for the unwavering support and patience of our parents.

DEDICATION

Greet to my family whom I love, and who help me with love, encouragement, and inspiration, whence this constant faith in the dignity of labor always takes its source. The belief in myself has helped me to overcome challenges.

ABSTRACT

In an attempt to reach the end product of rich sentiment categorization, this research offers an end-to-end sentiment analysis system which incorporates the state-of-the-art in deep learning methods in addition to traditional machine learning algorithms. Understanding of the public opinion has become critical in decision-making processes in areas such as politics, business and social observation due to the increased growth of user-generated information in places such as twitter. This study is based on Kaggle twitter sentiment dataset which was rigor legacy as data cleaning, normalization, tokenization, TF-IDF vectorization and class balancing using SMOTE to ensure reliability and reduces the biasness. This was caused by the demand for correctly and general-purpose systems of sentiment analysis. A variety models such as Multi-Layer Perceptron (MLP), deep learning model, interpretable machine learning classifiers such as; KNN, Decision Tree, Random Forest, Extra Trees (ETC) and advanced techniques employing gradient boosting algorithm such as XGB, LGBM and CatBoost with the help of the processed data were trained. The major novelty of this study is the designed integration of the two paradigms presenting the synergistic workings of neural architectures and ensemble methods in boosting for dealing with structured and capturing complex semantic relationships. While older models had offered efficiency and interpretable good models, the gradient boosting and ETC also regularly defeat other models with the ETC having the best accuracy of 99.12% as per a comparative assessment using the accuracy, precision, recall and F1, confusion matrices. The results obtained show that choosing the type of algorithm should be a tradeoff between processing resource, interpretability and complexity, and deep learning algorithms should be implemented in case of subtle identification of sentiment while simpler models are enough for getting quick insights. All thing said, this work allow to make repeatable pipeline with key focus on importance of preprocessing, feature engineering and balancing class for downstream success.

Keywords: ETC, Algorithm, Model, Accuracy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

The innovation of information and communication technology (ICT) over the last couple of decades has also played a significant role in the present day industries regarding customer service and experiences management. The digitalization, connectivity, and artificial intelligence are the transformations that customers have witnessed that enable businesses to connect with their values, preference, and behavior in a better way [1]. ICT has become a vital component of company policy and business performance in nearly all industries. The greater the customer satisfaction the better the image of the organization and the more competitive its business becomes in relation to revenue generation hence has made them to be a key constituent of the industrial sustainability and success [2]. In a recapitulation, businesses have gone a long way to interpret and assess responses largely via social media networking and web assessments to personalize their products to meet the dynamism of customer requirements.

The popularity of social media has enhanced the consumer engagement in the digital transformation field. Over 50 percent of the world population is accessing social media in some form or another, such as YouTube, Instagram, Facebook, and X (formerly twitter) [3]. These media are the places of immediate access to the sources of information, which are also used in marketing, communication, and entertainment. Online reviews is vast data set that has unstructured text data and it consists of subjective opinions, judgments and experience of the user with a product or service. Although it was stated that big data could be helpful in revealing the actions of consumers and the overall opinion, it was so large and complicated that a human mind could not analyze this large amount of data. Consequently, such massive sources of text data are becoming unavoidable and in case an individual would want to extract useful information out of them, the corresponding computing operations have to be automated.

Sentiment analysis, also known as opinion mining is one of the most commonly used methods of searching the textual content to find the subjective information [4]. It assists in defining the degree of polarity of the text: a neutral statement or opinion, a positive one, or a negative one and, accordingly, the tone and demeanor of the writer were emotional. Sentiment analysis is handy to identify the thoughts of the customers about a brand, product, or service. Social media post sentiment analysis, online reviews and customer feedback form sentiment analysis can contain valuable data useful in the customer relationship management and marketing strategies, or even product development. Devika et al. state that such a process is reflective of such an attitude or perspective of the speaker regarding a topic, and the most important benefit of the tool is that it can help companies obtain a real-time customer sentiment [4]. Determining the language signs or clues that can put different categories of feelings into right classification is a significant issue of sentiment analysis but not [5]. The human language is difficult to mechanically classify, as it is ambiguously constructed by nature and the expression of emotions might be subtle or even sardonic or even context-dependent.

There has been extensive academic and commercial research in the consumer sentiment analysis area due to the proliferation of ICT [2], [6]. A number of studies have focused on the capability of the system to utilize the feedback information provided through the social media postings or online reviews to identify the customer happiness and quality of service. The most important aspect of this initiative is integration of text mining and natural language processing (NLP) techniques that aims at attracting information within the computer through the unstructured text in the natural language [7]. The processes that are normally applied to the text mining pipeline to prepare the text data to this kind of categorization include tokenization, stopword filtering, stemmed words and vectorization. As Jain et al. [7] state, such preprocessing methods are central to the development of tools and methods to verify the integrity of data and, therefore, enhance the performance of the machine learning model.

A number of studies have studied the topic of review mining by utilizing machine learning (ML) methods [8], [9]. Positive, negative or neutral categorization of texts is commonly done using machine learning algorithms, including Naive Bayes, Support Vector machine

(SVM) and Logistic Regression [10]. The techniques are based on feature extraction algorithms that is a form of numerical representation of the text, i.e. bag of words (BoW) or term frequency inverse document frequency (TF-IDF). Although these techniques might be basic and effective, they do not assist individuals to understand sarcasm, context, and complicated linguistic interdependence. One of the first, but important, steps towards computerized interpretation of the customer perception and views lies in the field of data mining and text categorization relied on analyzing the consumer opinion, asserts Alhojely [10]. Sentiment analysis is also expounded by Pang and Lee [11], as a computer method of identifying and classifying the thoughts and perceptions of customers concerning a given product or service. To make data-driven decisions, social listening will enable the company to know, track and comprehend common opinions formed by social media and other online communications.

In this new digital age, consumer-business connections have already been altered radically. Previously, potential customers were very much dependent on personal contacts or word-of-mouth recommendation when making their shopping choices of purchasing products or services. But with the introduction of the internet reviews, ratings and discussion forums, these conventional methods are no longer relevant [12]. Customers get access to thousands of reviews of various products within few seconds that is faster and enhances the quality of their buying choice. Also, a sustained dialogue within the internet environment can help businesses to gain insight into consumer satisfaction and performance in the market as well as track the performance of their businesses on a regular basis without necessarily relying on the traditional survey or focus group methods. This has altered the method that corporate analytics and made it gather input to lead to constant response and enhancement of a service.

The review article has been discovered to have the potential of saving a significant amount of money and time when analyzed through machine learning methods [12]. With machine learning models, it is possible to test high volumes of data as compared to the case of analyzing high volumes of data manually with an associated time delay and probability of subjective bias. The opinions of customers can be collected based on the reviews of the site since machine learning algorithms are art to find the patterns and the relationship between words, phrases, and emotions and moods. Receipts are frequent with topic gaps (parts on products or services) on which the customer expresses his opinion, whether good or bad, on one aspect. Verma et al. argue [12] that the automation of the procedure also contributes to lessening the cost of operation and enhancing the quality of the decision because it offers an objective sentiment analysis.

The deep learning methods, in more recent times have been used to improve the performance of the sentiment analysis. Hung and Alias 5 demonstrated that DL based sentiment models did better than the ML classifier in complicated linguistic structure. The handcrafted features are compatible with traditional ML models, however, the DL models (RNN and CNN) have the potential to automatically derive an embedded textual representation with semantic and syntactic relationships to enable subtle sentiment predictions. True to this, Park et al. [13] asserted that aspect-based sentiment analysis can also be referred to as topic-based sentiment detection, which can provide detailed information on whether an individual product feature is liked or disliked, and polarity detection.

Sentiment analysis is a highly useful tool in the determination of the consumer sentiment in the present day data-driven environment. In these days machine learning is used in text mining through the aid of information and communication technology (ICT) and it has proven successful in terms of transforming large mass of waste unstructured data into knowledge to influence the strategic thinking. Among other things, such analytics can assist law firms to ensure that they better forecast the market and improve the satisfaction of the clients and the quality of the services. Those that are more complicated, such as context interpretation, sarcasm detection and multilingual sentiment comprehension, or multilingual sentiment categorization, are still the subject of further research despite the fact that machine and deep learning techniques development has progressed considerably. Sentiment analysis will play a bigger role in closing the gap that exists between artificial intelligence and human emotion as ICT evolves. It will also contribute to the improved knowledge of consumer psychology and the way people converse in cyberspace.

1.2 Motivation

This research is driven by both intellectual curiosity and personal motivation, as sentiment analysis is both a good research area that bridges many fields, including linguistics, psychology, computer science and data science. Its interdisciplinary character provides useful opportunities for the advancement of theoretical understanding as well as useful applications. The study aims to fill or reduce the existing gaps in the current literature by providing a refined methodological viewpoint on the efficiency of classified sentiment analysis achieved by comprising modern deep learning architectures and conventional machine learning methods. Such advancements have the potential to elevate user experience levels and to enhance academic debates about natural language understanding and broaden sentiments analysis technological advancements into the various sectors.

The driving factors behind this work are the increasing importance of opinion mining in the commercial, governmental, and societal spheres, the inherent complexity of human language, and the enormous possibilities that are arising with the advent of big data and the tools that are enabled by artificial intelligence. The main goal is to suggest methods that can be used to increase the accuracy, reliability and scalability of sentiment analysis systems. This study is inspired by both theory and practical need.

1.3 Objectives

i. To provide the comprehensive framework of sentiment analysis by combining current deep learning methods and classical machine learning models.
ii. To apply cleaning, normalization, tokenization, TF-IDF, and SMote in preprocessing Twitter Sentiment as provided by Kaggle, to have balanced and reliable modeling.
iii. To train and test various classifiers, e.g. Multi-Layer Neural Network, KNN, Decision Tree, random forests, Extra trees and gradient boosting classifiers (XGB, LGBM and CatBoost).
iv. To evaluate the performance of the model using the following measures confusion matrix, F1-score, accuracy, precision, recall and classification report.

v To experiment with trade-offs among interpretability, complexity and computing efficiency to choose the most appropriate sentiment classification model.

vi To develop a sentiment analysis pipeline that is reproducible and ubiquitous in terms of real-world applications and settings like customer feedback analysis and social media surveillance.

1.4 Rationale of the Study

The rationale behind the chosen study is that the increasing necessity to measure and examine more specifically the opinion of the population exists, at the moment when social media, as well as user-created materials, are on the center stage. The daily social media like twitter produce immense amounts of textual information that contains valuable information about the opinions of the population that can be utilized to guide corporate, political and social behavioral analysis decision processes. But there are major challenges in the traditional method of analysis on the unstructured and chaotic nature of this data. It is a well-built, trustworthy and generalized sentiment analysis system that can effectively operate and classify such data that is highly desired. It is an attempt to fill the gap between interpretability and prediction ability by concentrating on uniting the modern approaches to deep learning with the traditional machine learning algorithms. Other strategies like strict preprocessing, feature engineering and class balancing like SMote are also observed to achieve good performance of the model and minimize the bias. Such a strict procedure makes sure that eventually, the resulting framework will not just provide the most outstanding accuracy, but will be versatile enough to serve a vast number of purposes in the real-life.

1.5 Research Question

- Why is Twitter data used in this study?
- What type of data preprocessing steps were applied in the study?
- What is the main goal of this research?
- Which machine learning algorithms were used for sentiment classification?
- What is SMOTE, and why was it used?

- Which model achieved the highest accuracy in this study?
- What is the role of TF-IDF in text processing?
- Why is it important to compare different models in sentiment analysis?
- What are some real-world applications of this sentiment analysis framework?

1.6 Expected output

It is assumed that the research will give a high accuracy and reliable sentiment analysis framework capable of classifying the text data in Twitter to positive and negative sentiments. It is hoped that the combination of the benefits of both the conventional machine learning models and more modern deep learning methods will see the framework demonstrate improved performance, strength, and generalizability than the performance that is provided by either method alone. It is expected that the joint application of these steps such as cleaning, normalization, tokenizing, TF-IDF vectorization and class balance with SMOT will help to improve the quality of the input data and reduce the bias in the model to a minimum. Ensemble-based algorithms, as well as deep learning architecture, are likely to be more accurate, precise, recall and F1-scores. In general, the objective of the research is to develop a pipeline that could be reproduced with a successful transfer to the actual spheres, including social media monitoring, analysis of customer feedback, and market sentiment prediction.

1.7 Project Management and Finance

The project begins with a two months parallel work to determine the scope of the research project. Identification of the extensive issue procedure is aimed at establishing the fundamental objectives of the research in August. The in-depth Related Works and Gap Analysis will run in parallel to this and is within the same time frame. This concurrent timing is necessary in the successful launch of the project since it is through this that the research team will be made aware of what sort of unique contributions is being undertaken at the same time and the importance of the study will be solidified.

Once the theoretical basis has been fulfilled, the whole attention is focused on the production of actual data. The month of October is the only month when the data collection and preprocessing tasks are implemented. It should be within this stipulated time that all the necessary processes be completed beginning with the process of obtaining the raw text data (as in the example of product reviews) followed by its cleaning, normalization and tokenization and class imbalances management will lead to the modeling stage.

Basic stage of development is disintegrating in November. During this one month period (the process of building and training all the algorithms that will be used in the research including classic ml and deep learning models) is underway and is referred to as Model Implementation. This step will need hard, focussed work to build and sharpen the categorization systems, until one puts them through their paces.

The final stages of the project have three overlapping key jobs that will be involved in order to make sure that this project is completed within the scheduled time. Simultaneously with the becoming of the models, the Performance Evaluation and Comparison job starts in November and continues to the first part of December. This makes it possible to do iterative testing and improvement. The two-fold demands of documentation and review and submission are owed at a last minute rush in December, immediately after the main assessment. To adequately close the entire program at the close of the year, this overlap will ensure that the final report is compiled and presented at the same time when the final evaluation results are will be available.

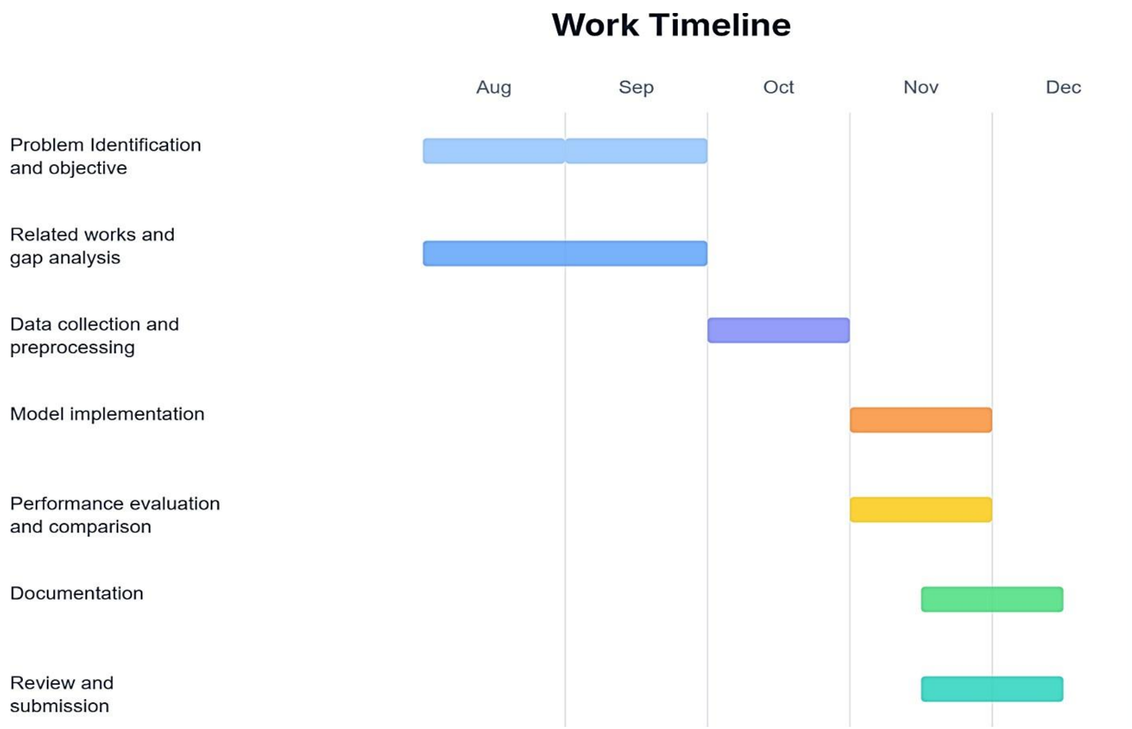


Figure 1.1 Work Timeline

1.8 Report Layout

This report is divided into numerous structured parts to help in a comprehensive understanding in the direction of trying to comprehend the sentiment analysis research. The introduction describes the background, description of the problem, justification, goals, importance and scope of the study. A brief overview of the sentiment analysis techniques is presented in the Literature Review, including the discussion of the current deep learning strategies, traditional Machine Learning techniques and findings of the earlier studies. The Methodology section includes the dataset used, data preprocessing techniques like cleaning, normalization, tokenization, TF-IDF vectorization, SMote based balancing and the description of the machine learning models, gradient boosting algorithms and the MLP deep learning model used, which is followed. Extra Trees Classifier is the most performing model, according to the Results and Analysis chapter that measures the performance of the models in relation to accuracy, precision, recall, F1-score, and confusion matrices. This is followed by the Discussion where the results are interpreted and compared with the other studies conducted in the past, pros and cons of the proposed framework are presented. The Conclusion and Recommendations, containing the primary findings of the study, its contributions, and recommendations on future developments follow the References and Appendices that are dedicated to the additional resources, i.e., codes, model parameters, and other figures.

CHAPTER 2

BACKGROUND STUDY

2.1 Preliminaries

Literature review provides the basic knowledge necessary in the exploration of how a successful sentiment analysis system can be created. Scientists have explored a number of strategies, both traditional machine learning models and advanced deep learning models, as sentiment analysis has evolved to become a necessary instrument in gauging opinions posted online. Earlier studies assert that language normalization, feature extraction, and preprocessing are essential steps in improving model performance. Although recent advances in gradient boosting and neural networks have demonstrated higher abilities to syntactically and semantically describe complex patterns in text, traditional machine learning methods such as KNN, Decision Trees, and Random Forest have also been used extensively because of their interpretability and efficiency. Moreover, much research has been conducted on how to deal with the issue of data imbalance, noise, and unpredictability in user-generated information, which makes it evident that the industry requires credible pipelines that will combine both conventional and modern approaches. This chapter puts the current work in the context of the wider frame of the research on sentiment analysis by examining the relevant theories, algorithms and empirical data.

2.2 Related Works

Sentiment analysis (SA), which is an alternative name of the so-called opinion mining, is an essential sub-field of natural language processing (NLP). It is focussed on "the design of computational systems with the ability to automatically interpret, classify and summarize emotions, attitudes, and opinions that are expressed with textual means." [14] With the advancement of social networking sites and online communication, this has led to the expression of people's thoughts, feelings and experience within the sphere of digital communication. Platforms such as blogs, micro blogs, review sites and discussion forums create a huge amount of written content, hence large opportunities for interpretation using data. However, the huge number of tokens and unstructured nature of these texts makes manual analysis unfeasible behind a growing need for automatic SA tools [15].

The exponential increase in information and communication technologies (ICT) has meant that social media is now inextricably a part of everyday interaction with billions of people active in the daily content creation of Facebook, Twitter, Instagram and WhatsApp. Such user-generated text not only describes individual viewpoints but have been identified as a significant source of information for businesses, decision-makers and researchers that are interested in knowing people's views about products, events and social issues [16]. While Hamad et al. [15] state the role of SA in the recognition of emotional patterns and customer's perspective, according to Gunasekaran [14], SA is an AI-driven technique that tends to automatically extract the emotional signals included in the text.

There have been a lot of developments in the application of machine learning (ML) and deep learning (DL) to the tasks in SA in recent years. These techniques have been seen to be more accurate, scalable, and context aware as compared with previous rule based or lexicon-based or dependent systems. Cui et al. [16] further highlights that SA has become one of the fastest growing area of NLP, which poses like multilingual processing, transferability across topics and sarcasm recognition. These advancements are also making it possible for organizations to take advantage of automated sentiment classification for decision-support, customer satisfaction monitoring, and integration with intelligent feedback systems.

Early SA models were built on traditional machine learning classifiers such as NB, SVM and KNN and textual representations such as Bag-of-Words (BoW), TF-IDF. Rajasekaran et al. [17], on the other hand, received 72.2% accuracy using SVM and NB of restaurant reviews which portrayed its efficiency to differentiate from the polarity in text. Similarly, Krishna et al. [18] demonstrated superiority of SVM in binary and multi class SA functions as well having promising accuracy of 94.56% in the classification of restaurant review. These results indicate that the classical ML models are able to find reliable patterns of the customers as well as to predict customer sentiments.

The integration of social media analytics has also been used to further increase SA applications. Yi and Liu [19] raise the problem that the posts from platforms such as Twitter and Instagram capture varied expressions of the public and opinions. This explosion of internet communication has provided an ideal arena for deployment of SA models. Jianqiang et al. [20] proposed a deep CNN for analyzing the Twitter streams and they were able to get a better prediction result of the sentiment by the feature fusion and feature extraction of local and global texts. Their approach demonstrates CNN's ability to take a big amount of data, thus to process it efficiently and to improve the classification accuracy.

Preprocessing is a major step in SA research to guarantee the reliability of the model. Techniques such as tokenization, stop word elimination, stemming, and lemmatization can be used to help clean and standardize the text before it is used to train the model. Given the informal and noisy nature of the structure of the tweet (including the hashtags, emoticons and abbreviations), Jianqiang et al. [20] stress the importance of a huge amount of preprocessing. Geler et al. [21] shows that NLP based artificial intelligence systems are able to detect positive, neutral and negative sentiments when the textual data is well prepared and available in the form of blog, comments, reviews and micro blogs.

The very commonly used analytical models in SA undertakings are those driven by ML. According to Adak et al. [22], these models are based on learning mechanism that results in adaptive improvement in accuracy of the prediction. Researchers have considered a number of algorithms including SVM, PCA and KNN to examine the performance in identifying the consumer's attitude. Verma et al. [23] reported that SVM had the highest accuracy (96%) since it has a great ability to handle the non-linear separations by kernel functions.

Gupta et al. [24] used Multinomial NB to dataset of Amazon, IMDB and yelp achieving accurate scores of 77% and 76% in a competitive rivalry with NB and SVM. In a similar line of research AlQahtani [25] compared BiLSTM, RF, NB and BERT on Amazon mobile reviews and proved that BiLSTM and BERT present the best performance results with accuracy of 94% and 90% respectively. These results provide an example of the superiority of DL models with a combination of contextual embeddings and sequential learning.

SA methods based on Twitter are mainly of three types: supervised, unsupervised and lexicon based methods [26-28]. In the case of supervised models the algorithms such as SVM, NB and logistic regression are trained on a text that is labeled. These approaches are usually equipped with contextual parameters such as POS tag, n-grams and emoticons in order to improve their robustness [20]. Lexicon-based approaches (e.g., SentiWordNet) grounded on predefined lists of words with their sentiment score; Developed by Baccianella et al. [29], SentiWordNet is a representation of the polarity values for words in the WordNet synsets. While fairly easy to interpret, lexicon strategies tend to get tested by managing context-sensitive data or domain-language and hence is preferred for data-based methods such as the ML and DL methods in more efficient ways.

BiLSTM is very important in the development of SA research in modern times. Due to its operation in both forward and reverse directions, it processes text and helps the model to capture entire contextual relationships as per the Rana et al. [30]. Ryan et al. [31] used BiLSTM to classify more than 56 thousand reviews of applications from the OTT to three negative, neutral and positive with an accuracy of 92%. Pratiwi et al. [32] used the combination of BiLSTM and Word2Vec-CBOW embeddings for Twitter reviews and achieved accuracy of 58.56% with 25 epochs - showing the sensitivity of BiLSTM models to data amount and quality.

A more advanced architecture by Cheng and Li [33] combined BiLSTM and pre-trained BER embeddings and attention mechanisms to better recognize multimodal sentiment recognition from text and audio feedback with accuracy of 85.3% on the CMU-MOSI dataset. This work reveals the potential of attention-enhanced BiLSTM systems for complex multimodal sentiment task.

Despite the rise of DL models, traditional ML classifiers like SVM and NB have their salt due to their interpretability, efficiency and present day theoretical power. Devika et al. [34] point out their reliability in the case that simplicity and speed of calculation are of the utmost importance. The efficacy of SVM in text classification in high-dimensions is repeated by studies by Nagelli et al. [35] and Alsemaree et al. [36]. Similarly, Wankhade et al. [37] state that both NB and SVM are still widely used as their implementation is very easy and they have a good generalization for different industries.

Over the years, SA has evolved from the level of lexicon-based systems to sophisticated ML and DL pipelines. Traditional ML methods (SVM, NB, RF, KNN) are still useful for structured dataset, DL models (CNN, LSTM, BiLSTM, BERT) are good for capturing semantics, context and sequence dependency. Yet challenges remain -- such as dealing with sarcasm, cultural nuances, noisy text and multilingual data. Future systems can be expected to use hybrid ML-DL and symbolic models to provide more in-depth and interpretable sentiment reasoning.

This project will focus on SA using ensemble methods including bagging, boosting, stacking and voting along with traditional ML and state-of-the-art DL models and evaluate the suitability of the models for sentiment classification tasks.

2.3 Comparative Analysis and Summary

The Comparative analysis of this study with existing proposed technology is compared below table 2.1.

Table 2.1. Comparative analysis with previous work

Authors	**Techniques / Models Applied**	**Reported Performance**	**Key Contributions / Advantages**	**Identified Gaps / Drawbacks**
Gunasekaran (2023)	Conceptual Sentiment Analysis Framework	Lexicon-driven semantic scoring: 88%	Offers a straightforward explanation of sentiment analysis and discusses how automation can support decision-making.	The study does not present real-world validation or practical implementation results.
Hamad et al. (2021)	Naïve Bayes Classifier	73% accuracy	Emphasizes the value of examining social media interactions for understanding public sentiment.	Fails to compare the classifier against other standard models.

Cui et al. (2023)	Survey of Modern NLP and SA Approaches	Not specified	Provides a broad overview of recent tools and research directions in sentiment analysis.	Remains theoretical with no dataset-based experimentation.
Rajasekaran et al. (2019)	SVM and Naïve Bayes	72.20%	Investigates real-world restaurant review datasets to evaluate classification performance.	Accuracy remains modest; lacks comparison with contemporary deep-learning methods.
Krishna et al. (2019)	Support Vector Machine	94.56%	Demonstrates strong classification potential and high accuracy using SVM.	Focused on a single dataset, limiting general applicability.
Yi & Liu (2020)	CNN-based Sentiment Model	98%	Introduces convolutional neural networks for processing Twitter sentiments with exceptional accuracy.	Heavy preprocessing is required; model struggles with context-rich or nuanced expressions.
Jianqiang et al. (2018)	Deep Convolutional Neural Network	81.36%	Combines multiple features within a deep CNN to enhance prediction capability.	Computational demands are high, and interpretability remains limited.
Geler et al. (2021)	SMO, Random Forest, Random Tree, REPT, M5P, MP	MAE = 0.541	Highlights the ability of machine learning models to classify sentiment polarity across categories.	Findings remain abstract and lack detailed model comparison metrics.
Adak et al. (2022)	Machine Learning & Deep Learning Opinion Models	Lexicon accuracy: 87.33%	Outlines automated opinion mining methods and emphasizes ML applications.	Experimental verification is minimal or absent.
Verma et al. (2023)	KNN, PCA, SVM	SVM: 96%	Shows the superior effectiveness of SVM and discusses performance across multiple algorithms.	Dataset used is domain-specific, which limits model generalization.

Gupta et al. (2023)	Naïve Bayes, SVM, KNN, Random Forest	Random Forest: 78%	Conducts comparisons using Amazon, IMDB, and Yelp reviews to examine classifier performance.	Reviews without ratings were excluded, which may introduce bias.
AlQahtani (2021)	NB, RF, Bi-LSTM, BERT with Logistic Regression	BERT: 98.4%	Presents an extensive comparison of ML and DL models, highlighting the superiority of BERT and Bi-LSTM.	High resource requirements and limited model transparency.
da Silva et al. (2014)	MNB, SVM, Random Forest, Logistic Regression	76.25%	One of the earlier works that structured sentiment analysis as a machine learning task.	Considered outdated when compared to modern transformer-based techniques.
Hagen et al. (2015)	Supervised ML Techniques	F1-score: 71.09%	Utilizes POS and contextual signals to analyze Twitter sentiments effectively.	Restricted to tweets and lacks integration with deep architectures.
Saif et al. (2012)	Lexicon-Based + Supervised Methods	75.95%	Introduces early feature-enhanced lexicon approaches for sentiment prediction.	Lexicon size and scope limit accuracy.
Montejo-Ráez et al. (2014)	Lexicon-Driven SA	75.95%	Proposes techniques involving SentiWordNet for weighting sentiment.	Unable to detect sarcasm, figurative language, or hidden sentiment cues.
Paltoglou & Thelwall (2012)	Statistical + Lexicon Weighting	71.09%	Improves polarity classification by incorporating refined lexical weights.	Performs poorly on noisy or informal social media text.
Thelwall et al. (2012)	Lexicon-Based + Hybrid Methods	76.25%	Provides foundational approaches for automated sentiment scoring across datasets.	Considered outdated relative to modern deep-learning-based innovations.
Baccianella et al. (2010)	SentiWordNet Lexicon	Not reported	Supplies a widely used sentiment lexicon that supports numerous subsequent studies.	Static lexicon cannot accommodate evolving vocabulary or contextual variations.

Rana et al. (2024)	BiLSTM Architecture	92.33%	Offers a conceptual explanation of BiLSTM and its role in modeling sequential text.	No practical experiments or evaluations included.
Ryan et al. (2024)	BiLSTM, LSTM, CNN, Logistic Regression, XGBoost	92%	Effectively models app review sentiment using multiple architectures with strong accuracy.	Study is limited to OTT application reviews.
Pratiwi et al. (2024)	BiLSTM with Word2Vec (CBOW)	58.56%	Uses embedding-driven sequence modeling for capturing contextual sentiment.	Moderate performance suggests need for more data or model enhancements.
Cheng & Li (2024)	BiLSTM + Attention + BERT + Audio Fusion	85.30%	Introduces multimodal sentiment analysis combining text and audio signals.	Complex framework requiring substantial computational power.
Nagelli et al. (2025); Alsemaree et al. (2024); Wankhade et al. (2022)	SVM, NB, DT, CNN, LSTM, RNN	88%	Summarizes findings showing classic ML models like SVM and NB remain effective in several SA tasks.	May struggle with generalizing to unseen or highly nuanced text scenarios.

2.4 Scope of the Problem

The analyzed literature presents a number of gaps and issues according to which the problem of sentiment analysis study is characterized. Even though several studies have been carried out on lexicon-based, machine learning, and deep learning methods, their effectiveness is highly varied based on datasets, domains, and preprocessing methods. The older algorithms like Naive Bayes, SVM, and Random Forest tend to make middle-level accuracy but they cannot do generalization and situational interpretation and lexicon-based approaches are constrained by the static dictionary that cannot pick sarcasm, slang or implicit emotion. Deeper learning algorithms (like CNNs, LSTMs, and BiLSTMs) have better accuracy in most of the cases, but have high computational requirements, intricate preprocessing, and large and heterogeneous datasets, not always available in a study. Some of the works are not empirically tested, compared with some benchmarks, or validated in real-life, which decreases their practicality. Also, the lack of stability in the performance of domains of restaurant reviews, product reviews, and Twitter data indicates that creating a sentiment analysis model that works effectively in all settings is challenging. All these difficulties add to the necessity of a unified, thoroughly validated system that incorporates preprocessing and feature engineering, along with machine and deep learning, spectrophotometry to produce a reliable and generalizable sentiment classification.

2.5 Challenges

- **Data balancing:** It has been observed that a number of problems occur during the training and the overall performance of the models during the result collection stage. One of the major problems was the data imbalance, when certain classes of sentiment (such as good reviews) were overrepresented, as compared to others (such as negative or neutral reviews). This gave biased predictions and less accuracy in minority classes, but the first choice of the model is the majority class. Although the synthetic oversampling was balanced with the use of SMOTE, occasionally a noise would be generated and this to some extent affected the accuracy.
- **Feature Extraction:** Extraction of features was another challenge in this field using a mixture of traditional and emerging techniques. Bag-of-words and TF-IDF models are sparse matrices of higher computational complexity and required a lot of memory and processing as well as training was slow. Embedding models such as BERT require vast amounts of memory and processing as well as training is slow. A major setback was the computing constraints and the ways through which these models could be ensured to be implemented effectively.
- **Data Cleaning:** Data cleaning was also challenging because outliers such as emojis, slangs, and misspellings tend to be part of product evaluations. Normalization and lemmatization required multiple cycles before the correct findings were obtained and a sufficient amount of information was gathered, not

compromising the contextual meaning. Also, it was not easy to use preprocessed tokens to obtain semantically consistent embeddings.

- **Model Training and evaluation:** The evaluation of the model demonstrated that it is not stable on different runs especially in instances of hyperparameter adjustment. The ultimate compromise between the different metrics of accuracy, precision, recall, and F1-score took a long time to come up with. These difficulties notwithstanding, the results were later brought under control and attributed to a sequence of progressive advancements in preprocessing, balancing, and fine-tuning of the models.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The current study used several open source text-based datasets largely used in the analysis of patterns of sentiments and emotions on online datasets. These datasets - Sentiment140 dataset, IMDb movie reviews dataset, Amazon product reviews dataset and Google's GoEmotions dataset - have been handpicked for this purpose since they have rich text with diverse and emotional nature suitable for both sentiment detection and fine-grained emotion classification. They include short informal tweets, long form written reviews and structured user comments tagged with categories such as positive, negative, neutral and more complex emotions such as anger, joy, sadness, surprise and fear. This combination of different writing styles and expressions of emotion give one a whole continent, from which it is possible to compare one computational model with another.

In order to process and analyse the data, different text cleaning and feature generation tools were used in the study. Python based NLP frameworks like NLTK, spaCy and regular expression patterns were used for tokenization, stop word removal, normalization, punctuation stripping and lemmatization. For the purpose of feature representation both classic vectorization techniques (Bag of Words and TF-IDF) and advanced embedding techniques (Word2Vec and pre-trained BERT embeddings) have been used. These representations were passed in to different machine learning algorithms i.e. Naive Bayes, Logistic Regression, Support Vector Machines as well CNN, ANN, LGBM, CatBoost and self-normalizing neural networks. Model performance was measured in the accuracy, Precision, recall, F1 score and also confusion matrices using libraries like Scikit-learn, TensorFlow, PyTorch etc. Combined these datasets and analytical tools formed a strong scaffold for the evaluation of the efficiency of various sentiment and emotion classification methods.

3.2 Data Collection Procedure

For the purposes of this research, a set of Twitter sentiment data available publicly was downloaded from Kaggle (Twitter Sentiment Analysis, 2021) [38], a quite popular open datasets and data science competitions platform. This dataset was selected because of its great range of topical coverage, the up-to-date state, well-defined sentiment label which is very appropriate to train and test the sentiment analysis models. It consists of 73,812 tweets that are categorized in 4 types - Negative, Neutral, Positive and Irrelevant. In particular it happens to have 12,842 useless posts, 20,614 positive posts, 18,046 neutral posts and 22,310 negative posts. The distribution provides a fairly balanced structure to the comparison of different classes of sentiment. The data set that was available in excel can be found and was imported in python using the pandas and numpy module. Initial exploratory data analysis (EDA) was conducted to see if there was class imbalance and text length variation and common linguistic patterns. Subsequent preprocessing were carried out, including removal of hyperlinks, user tags, hashtags, emojis, special symbols; Lower case conversion, and tokenization. Although the dataset was already labeled first, additional cleaning had to be done in order to clean the input to standardize it before training the model. Notably introduction of "Irrelevant" class helps in allowing meaningful emotional content or off topic/ context free posts to be differentiated by the models which helps in improving strength and performance of the model in real world. The usage of informal language, abbreviations, and spontaneous expression of emotions in this dataset is very similar to the way in which people communicate in Twitter, which makes it a good resource in the development and evaluation of machine learning and deep learning models related to detecting sentiment on social media.

3.3 Statistical Analysis

The statistical aspect of the research work was done with a detailed comparison of the performance of different models based on well defined evaluation metrics so as to ensure that it is reliable, valid and fair across algorithms. After conducting preprocessing techniques, the data was split into two sets (train and test) and the class imbalance problem was overcome by applying the SMote technique. Each model was evaluated in terms of terms of accuracy, precision, recall, F1-score and confusion matrix analysis to understand how well the models were able to classify different sentiment labels and if there were any repeating misclassification patterns. These different metrics helped to quantify the ability of each algorithm to tell-apart contrasting emotional tones. Comparative results showed that ensemble-based approaches were more accurate than traditional classifiers, boosting models, and the architecture MLP. Overall, the structured evaluation process made it possible to find the best statistical model in terms of consistency and robustness to classify the sentiment.

3.4 Proposed Methodology

Data Pre-Processing

The dataset went through a number of preparation steps to make sure that it was clean, consistent and ready for machine learning and deep learning models. Rows that had empty text fields were filtered out and all other entries were changed to the string format in order to make them easily handled. The text was enhanced by removing URLs and hypertext markup language (html) tags, symbols and characters that were not letters and then by converting text to lowercase. After being cleaned, the text was tokenized, and a TF-IDF vectorizer was applied in order to transform the tokens into numerical features. SMote was used to overcome the issue of imbalance of dataset and dataset was segregated into training dataset with 80% data and a testing dataset with 20% data. The resulting sparse TF-IDF matrix was then transcended to a dense format in order to be compatible with some algorithms.

Removal of Missing Values

Taking care of missing data is a key thing, as incomplete text inputs break NLP processes and can lead to the faulty results of the model. Null or empty values in the Text column will lower the robustness of patterns derived from the data and also may cause technical problems while performing tokenization or embedding. Removing these rows helps you to keep the data set structurally sound and will help you to avoid any errors in your modeling that will occur later on.

Data Type Conversion

To achieve consistency, it was converted key columns to the right data types. Text values were forced into string format in order to be used for NLP operations and sentiment labels were turned into categorical or numerical values that are needed for machine learning algorithms. Proper type conversion assists in staying away from processing mistakes and making the memory utilization process for training the model efficient.

Data Cleaning

Raw user generated text often has extraneous noise. Cleaning steps which included removal of links, html code, emoji's and punctuations were performed to make the text easier to process and for the model to be more accurate. The way this works is taking messy and informal language and presenting it in a more uniform way such that it's readable by algorithms in a reliable manner.

Text Tokenizer

Tokenization refers to dividing the text into smaller pieces that can be analyzed by models. Since tweets are prone to slang words, hashtags, abbreviations, state of the art tokenizers from libraries like NLTK or spaCy were used instead of the simple whitespace methods. Proper tokenization helps to lay the groundwork to embed algorithms such as TF-IDF or Word2Vec and also prevents the problem of inconsistent input lengths when using deep learning algorithms.

TF-IDF Vectorizer

TF-IDF converts the words into numerical values that have been weighted based on the frequency of their occurrence in a document relative to a document dataset. This helps accentuate key and important terms that have tends to have particular sentiments and reduce ordinary words that are basically functions as fillers. The method is widely used for big text data sets and prepare text in order to use in machine learning algorithms.

Label Encoder

Sentiment categories, such as Positive, Negative, Neutral and Irrelevant were translated to numeric values so that they can be processed by the models. Encoding these labels to make the standardize format of the output and also for the compatibility of the supervised learning algorithms that work on numerical data.

Data Balancing

The dataset was non-symmetrical in the number of samples for the classes of sentiment. To ensure that models do not get biased against minority classes, SMote was used to oversample the minority-classes. The balancing of the data leads to more equitable reliable predictions of all sorts of sentiments.

Train-Test Split

The data set was split into two as a training data set and a test data set so the model can learn from one data set and be tested with the other unseen data set. An 80:20 split for detection of overfitting which would also ensure the performance measures like precision, accuracy, recall and F1-score would be reflecting the usability in real life.

Matrix Conversion

TF-IDF produces sparse matrices; on the other hand some algorithms need dense numerical matrices. Therefore the sparse representations have been transformed to dense matrices which makes them fully compatible with models based on dense computations. This step is in order to ensure a smooth training and evaluation of the model.

Used Algorithms

This sentiment analysis research involved a slew of machine learning and/or deep learning techniques that identify the degree of accuracy with which different machine learning models can interpret emotional tone based on written information. Logistic Regression served as the baseline model because of its clarity and simplicity in the computations as well as its suitability in binary and multi-class classification tasks. It estimates the likelihood of the class membership by a linear decision boundary and works reliably when are the predictors linearly separable. Along with this, another key reason of choosing Random Forest Classifier was the fact that it is an ensemble based method and can capture the non linear interaction by creating so mafic such independent decision trees and aggregating the results, which results in reduction of the model variance and makes the prediction more reliable. For comparison a single tree also was incorporated, Decision tree Classifier, which provides an intuitive easy to visualize structure and may go to overfit easily without using proper pruning.

In an attempt of investigating advanced boosting frameworks XGBClassifier and LGBMClassifier were implemented. Both are based on the construction of trees sequentially, where each successive tree solves the remaining errors from the previous one, with better accuracy, particularly with large-size datasets. LightGBM uses histogram based splitting for faster computation and light memory load and XGBoost is very famous for its regularization techniques and scalability and high competitive performance. Another gradient boosting model, CatBoostClassifier, was also used because of its robustness with regard to categorical variables and built-in mechanisms of reducing overfitting, which works perfectly for high-dimensional textual representations.

Furthermore, we included the Extra Trees Classifier also as another ensemble learning technique which differs from the Random Forest in the more randomised approach to splitting the trees, which more often than not results in increased generalisation capabilities. KNeighborsClassifier, a simple and powerful classifier that is non-parametric in nature, was also experimented with which enhanced the method of assigning the sentiments labels by looking at the closest neighboring samples, although in such a case, as the database grows, the computation time also starts increasing.

Two types of deep learning framework i.e. Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) were used also to capture the subtle variations in the patterns in the text. ANNs consist of interconnected layers of neurons that can learn hierarchical, non-linear relationships in the feature space. CNNs, which were originally created for image recognition, have been applied in the text analysis scenario, in which sentences are treated as sequences and local semantic patterns are extracted via convolution filters. Additionally, the MLPClassifier type of feedforward neural networks with multiple hidden layers was used, since they can process the TF-IDF feature vectors and classify the data with complex nonlinear relationships.

Logistic Regression

Logistic Regression is a classifier method and is basic in nature which is used widely for binary and multi-class problems. It predicts the probabilities of a piece of input in a specific category as well as maps out outputs in the probability space by a sigmoid function. Due to its linear nature, the model is very simple to interpret, as well as with a low computational requirement. Logistic Regression is very efficient in situations where there is more or less a linear relationship between the features and labels and is often something much more elaborate algorithms are compared against.

Random Forest Classifier

The Random Forest Classifier is an instance of ensemble learning which creates large number of independent decision trees and combines their results to give an output. This averaging mechanism has a committed drawback of decreasing the overfitting and boosting the generalization. Random Forest works very well with high dimensional data sets, can model non-linear interactions and it can naturally give a ranking of the importance of features. Its robustness, scalability and adaptability make it a reliable choice for text classification tasks, including sentiment analysis.

Decision Tree Classifier

The Decision Tree Classifier is based on hierarchical and tree structured model which divides the dataset in subsets based on some features. This structure is similar to a flowchart and therefore has become very interpretable for the user and fitting for identifying complex relationships for data. However, decision trees are susceptible to overfitting when not constrained properly in that it is possible for very small differences in the data to results in whole different tree-structures. Despite that, decision trees are the backbone of a lot of advanced ensemble algorithms like graduates boosting and random forests.

Extreme Gradient Boosting (XGB Classifier)

Extreme Gradient Boosting (XGBoost) is a powerful gradient boosting algorithm which is known to have a high level of accuracy, speed and efficiency. It constructs the decision trees in an ordered fashion and every tree attempts to minimize the errors of the previous tree. XGBoost has regularization terms which prevents overfitting and makes it a great choice for dealing with large, complex datasets. Because of its scalability and optimised processing in parallel with all the above-mentioned technologies, it is always performing on the top in different machine learning challenges.

A related model is the Extra Trees Classifier which adds randomness to the tree building process by selecting for random split thresholds rather than optimal split thresholds. This can add a greater diversity to the ensemble and can often increase the predictive performance and reduce the variance. Extra Tree is known to be able to handle a high-dimensional text data efficiently and is also known to be very fast and reliable.

Multi-Layer Perceptron (MLP Classifier)

The Multi-Layer Perceptron (MLPClassifier): is a feedforward neural network with one or more hidden layers which is trained with backpropagation. MLP is capable of learning very nonlinear patterns, hence when the numerical representation is RF-IDF vectors for example, MLPs are useful here in sentiment classification. Whilst less computationally efficient than traditional machine learning models, MLPs can often have good accuracy on structured and semi-structured text features provided they are properly optimized.

K-Neighbors Classifier

The K-Neighbors Classifier (KNN) is a very intuitive, instance-based, learning algorithm that provides class labels using majority sentiment of the nearest neighboring samples. Since it is not making any specific assumption on the distribution of the data, it is flexible and easy to implement. However, KNN could be a memory intensive and also slow in terms of the time taken to make a prediction as the size of the data increases. For all such shortcomings, it works pretty well when decision boundaries are non-linear and there is a high nonlinearity between features.

Light Gradient Boosting Machine (LGBM Classifier)

The Light Gradient Boosting Machine (LGBMClassifier) is an efficient and extremely fast gradit boost algorithm to the optimal-speed/low-memory ratio. The growing of trees is done in a leaf-wise fashion and deep accurate trees will grow given enough available data. LightGBM is especially suited to massive and high dimensional data making it a great choice for sentiment analysis tasks involving the use of many text-based features.

CatBoost Classifier

CatBoost Classifier is another type of gradient boosting model that has been designed with an aspect to deal with categorical features automatically, without any elaborate preprocessing or manual encoding. It includes ordered boosting and permutation based approach to overcome overfitting and to improve the performance. CatBoost tends to give good results with little parameter tuning and is suitable for the wide range of text classification problem where one deal with mixed or high-dimensional features.

Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is inspired by the biological structure of neurons in human beings. It consists of inter-connected layers which can learn abstract and complicated patterns by an iterative process of training. ANNs are good at modelling nonlinear relationships and are well-suited to modelling different types of data, and consequently are well-suited for natural language tasks when combining them with numerical text representations such as embeddings or TF-IDF features. Although ANNs can provide powerful predictive accuracy, they require a lot of computational power and the tuning of hyperparameters.

Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs), whilst originally designed for handling images, have been very successful in text classification. They work by using convolutional filters over sequences of words which allow the model to detect local dependencies, i.e. n-grams or phrase patterns. CNNs learn hierarchical representations of features and they are often high performing as compared to traditional models when the text is unstructured or noisy.

Their efficiency, high generalization capabilities and capacity of investing relevant semantic cues make them valuable tools in the sentiment analysis field.

3.5 Proposed Methodology

Sentiment analysis, both machine learning and deep learning techniques, is of a series of systematic processes, from data preprocessing, to all the way to model performance evaluation. To do this the data set which is used for the research is collected from Kaggle and contains Twitter Sentiment data which is first subject to a stringent cleaning process to get accuracy and consistency. This means that first all the missing or null values of text column are removed to avoid inconsistencies during the analysis. After the process, all the textual information is changed to a common format of strings, to make the information standard. The next step includes making all the characters lower-cased and removing any other extraordinary stuff like URLs, html tags, special symbols and other unnecessary noises in dataset.

For instance, in order to make the textual content amenable for computational modeling/handling, a process called tokenization is applied, that is split the cleansed sentences into individual units, or tokens. These tokens are then converted into numerical vectors with the help of the vectorization technique called the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. This method creates sparse matrix representation of dataset where each word is given some weight based on the frequency of word in a document with respect to its frequency in whole corpus of text thus showing words which has some important information of sentiments.

In order to make the sentiment labels work with machine learning algorithms, label encoding is used, in which the different classifications of Negative, Neutral, Positive and Irrelevant are assigned to numbers. Since in real-world scenario, the datasets have many times an unbalanced class distribution so for that Synthetic Minority Over-Sampling Technique (SMOTE) is used to create the more samples of underrepresented classes. This procedure help to ensure that the training data is balanced and that we are not training the model to be biased towards the majority classes, which can help to improve the model's ability to generalize.

Once the dataset is balanced the dataset is split into training and testing dataset in 80-20 split for being able to test the model performance with previously unseen data. Sparse TF-IDF matrices are then transformed to dense arrays for implementing deep learning algorithms, which need the dense numerical form of input.

A massive range of algorithms are implemented in the comparison of traditional machine learning and state-of-the-art deep learning algorithms. Gradient boosting models like XGBClassifier and LGBMClassifier are used for its efficiency of structured data sets whereas robust ensemble methods like Random Forest Classifier and Extra Trees Classifier are tested for its utility of decreasing variance and enhancing predictive accuracy. Decision Tree Classifier-The Decision Tree Classifier is used for its transparent and interpretive decisions, which is a form of hierarchical decision rules. Instance based algorithms like K-Nearest Neighbors (KNN)-categorize the data points based on how close are they the data points in the training data set, CatBoost Classifier - It is a type of gradient boosting algorithm, and it is particularly good in handling categorical variables. Multi-Layer Perceptron (MLPClassifier) is a feedforward neural network, which has also been used because of its ability to learn complex nonlinear patterns which makes it suitable for deep learning applications.

Furthermore, in order to evaluate the model performance, several quantitative metrics such as accuracy, precision, recall and F1-score are used. Additionally, confusion matrices enable a detailed analysis of the model's performance, considering individual categories of sentiment, which makes misclassifications patterns very clear. This end-to-end pipeline from preprocessing of raw data, creating the model based on the generated data, to evaluate the performance of the model helps to ensure that they are effectively used to capture and classify the sentiment in the text embedded in the data.

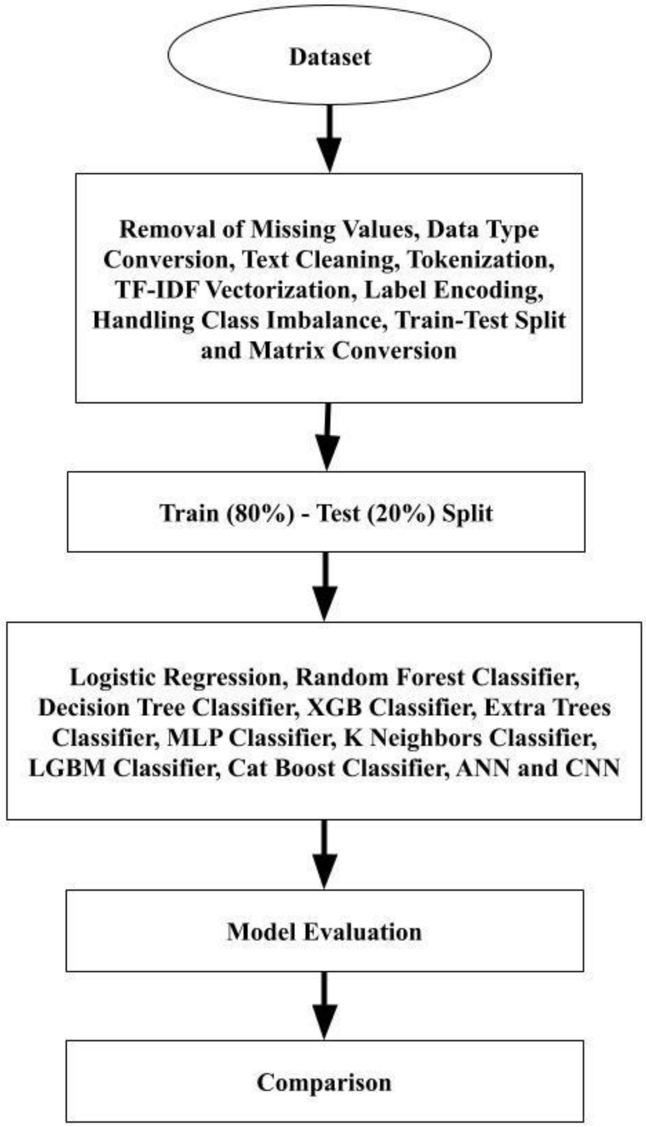


Figure 3.1 Flow diagram of the methodology

Evaluation Matrices

In machine learning and artificial intelligence, the process of determining how well a model is doing is just as important as building the model. Without a proper evaluation, it is impossible to know whether a model is reliable, general of data, or appropriate for the purpose for which model has been developed. Classification models are measured against various metrics and based on the metric it is measured from, a different perspective of the effectiveness of a model is measured. Commonly used measure will be accuracy, precision, recall, F1-score, confusion matrix and classification report is. Taken in combination these metrics give a more complete picture of how a model is going than one piece of numerical information. Accuracy is one of the most intuitive values which is the measure of how much is the correctly predicted number of instances to the total number of predictions. While it is simple and we all know how it works, it can be misleading in datasets which have class imbalance where there is one category which is more dominant. In such cases, while the overall accuracy of a model may be high, the model may not be able to accurately predict minority classes, which may be important in certain cases such as disease diagnosis or fraud detection. To overcome this shortcoming, the precision and recall has been used in which it gives a more detailed picture of the classification performances. Precision is evaluation of the ability of the model to correctly predict the positives and determines percentage of those prediction were correct are true positives predicted positive of total of things that were predicted positive by the model It is especially important in cases where false-positives have grave consequences such as when spam is being detected or rare diseases being screening for. Recall is also known as sensitivity or the true positive rate; recall tells the percentage of true positive instances that were correctly identified by the model. High recall is important in applications where failing to detect a positive case might have severe consequences (such as medical testing, or monitoring for fraud). The F1-score is what takes a mixture of the precision and recall and sort of gets them all together with a single metric using the harmonic mean which gets sort of a fair evaluation. This is a particularly useful score in datasets that have skewed class distributions in order to provide a balance between precision and recall and identify the effectiveness of the model in cases where accuracy alone may be misleading.

The confusion matrix is a breakdown of how a model is predicting the correct value as compared to the actual labels. It displays the number of BP's TN's FP's FN's that practitioners can identify areas of misclassifications numerically. This tool is instrumental in diagnosis weak points in a model and to have a sense of the types of errors it makes. The classification report group of a number of measures including precision, recall, F1-score and support for each class to provide an overall view which is easy to compare between different models or categories. Using multiple evaluation measures can help to get a more realistic understanding of the model performance in addition to a measure of accuracy especially when working with imbalanced data.

Accuracy

Accuracy is one of the commonly known and used evaluation metrics of data science and machine learning. It is the measure of the percentage of predictions that a model is able to get right out of the total number of predictions. This is an easy to understand measure of model performance, and often will be the first statistic that is used by analysts in evaluating a model. High accuracy may be used to suggest that the model can accurately predict the results in most of the cases, which can be considered as preliminary confirmation of effective model behavior.

Precision

Precision is an important metric in classification problem or certain problem where false positive is particularly expensive or if cost of failing to properly identify positive samples exceeds accuracy of the total result. It is the ratio between the number of true positive prediction of all the prediction that is made as positive. This metric is very important in the real world in which false alarms can have serious consequences (e.g. unnecessary medical treatment, misdiagnosis of disease or classification error in emails). High precision means that positive predictions can be surety in the prediction and there is least chance of false alerts.

Recall

Recall Is additionally talked aboutas sensitivity, recall is a measure of a model's capacity to acknowledge all relevant positive intent on a dataset inside. It figures out the ratio of actual positives than were called true positives to actual number of positives. A model with low recall is unable to get an important part of the positive cases while the high recall model manages to catch a majority of relevant cases. Recall is very important with cases like medical diagnosis, fraud detection and safety monitoring, where the failure of detecting a positive instance may have serious and even fatal consequences. In some high-stakes circumstances maximisation of the recall might be the most important consideration even over against other issues like accuracy.

F1-Score

The F1-score is a common method of determining metrics that is a balance between the two metrics (precision and recall) into a single metric that represents a performance. It is especially useful in cases of skewed datasets, and the metrics of precision and recall are not sufficient for indicating the model's performance. The F1-score is a combination which would express the ability of the model to form a balance between false positive and correct positive identification. It gives more weights to lower values and protect from the problem that giving higher score to lower values due to fine difference in between precision and recall. This makes it an important measurement in applications where the cost of the 'false positive' result along with the 'false negative' result is great.

Confusion Matrix

The confusion matrix is a statistical tool which is used to assess the performance of classification by comparing the predicted values with actual. As opposed to having a single accuracy score, it gives a detailed picture about how a model classifies amongst multiple classes. In a binary classification problem, the confusion matrix has four major elements -- true positive, true negatives, false positive and false negatives. This breakdown is enabling practitioners to know not only how often is a model correctly predicting, but to know what kind of error a model is making, important in tuning model performance.

Classification Report

The classification report combines a number of evaluation metrics together with a comprehensive summary of the results such as; precision, recall, F1 score, support to the summary for the classes. It is helpful in getting an insight into the model of performance, particularly in cases where classes are imbalanced. And unlike accuracy, which is potentially misleading with unbalanced datasets, classification report takes a class by class, which opens the door to something that finally, opens up a window that shows where the model is good or bad. This makes it an essential tool for understanding the nuanced performance of classification algorithms, and for providing guidance on how to make improvements in the models.

3.6 Implementation Requirements

Developing this sentiment analysis system took a combination of software tools, programming environments and computational resources in order to allow the data cleansing, training of the models and evaluation of model performance. Python was utilised as the main language because of its huge Library like NLTk and spacy for text cleaning and tokenization ML, pandas and NumPy for Data Manipulation and Scikit-learn for Using TF-IDF Vectorization Process and Machine Learning Algorithms. Deep learning model like Multi Layer Perceives (MLP) using TensorFlow or Keras and specializing version of boosting model like XGBoost, LightGBm and Cat Boost using implementation.

Visualization and evaluation of model metrics was done possible by Matplotlib and Seaborn which helped in the production of performance plots and confusion matrices. While classical models of machine learning could be run on regular CPU setups, in order to accelerate the computational process of deep learning models, computers with adequate RAM and GPUs dedicated to this purpose were needed. The fact that the upward reception of the sentiment analysis pipeline was successful was due to a good structure in the codebase, stable environment in the software environment, and good computational resources for every step, from the preprocessing to the model evaluation would be successful and efficient.

CHAPTER 4

Experimental results and discussion

4.1 Experimental Results & Analysis

In this investigation a great number of machine learning techniques and deep learning techniques which have different advantages in a classification task were used to perform the sentiment analysis. Foundational statistical methods, such as Logistic Regression (LR), were considered a good baseline, because of their simplicity and interpretability, whereas ensemble methods, such as Random Forest (RF), Decision Tree (DT) and Extra Trees Classifier (ETC), supported the use of numerous learners for predictive accuracy and to produce a more stable model. Sophisticated gradient boosting algorithms including XGBoost (XGB) and CatBoost algorithms as they have the ability to model complex data patterns and are able to appropriately balance classes and achieve higher performance on structured data.

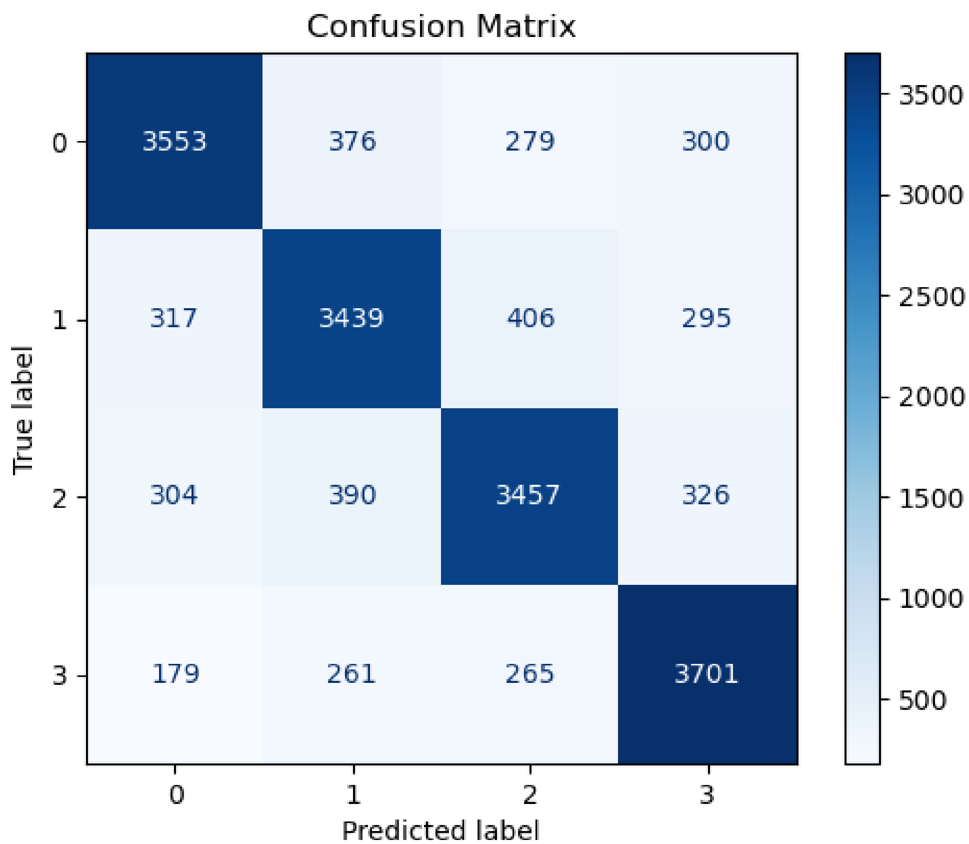
The neural network approaches were also included in the study which included Multi-Layer Perceptron (MLP) and Artificial Neural Network (ANN) which were capable of capturing very non-linear relationship in the data. Besides, K-Nearest Neighbor (KNN) instance based learning was used to build and make sentiment prediction possible based on similarity of data points. Convolutional Neural Networks (CNN) was further included, in order to capture spatial and contextual patterns in the textual features which helps to extract meaningful sentiment representations from the sequence of words.

By assessing the performance level of these different algorithms, the research was in a position to perform comprehensive comparison based on important metrics such as accuracy, precision, recall, and F1 Score. This evaluation provided crucial information about how well each model fits the problem of class feeling classification, how well they predict and how flexible they are to adapt to different sizes of data, and how well they can be used in practice in real-world tasks. The fusion of deep learning approach to traditional machine learning paradigms demonstrated the holistic approach to sentiment analysis by providing a well-balanced approach towards the machine learning algorithmic efficiency and feasibility for the deployment in real-world sentiment-driven tasks.

Logistic Regression

The Logistic Regression model gained 79.28% Accuracy, Precision, Recall and the F1 Score was 79.26% This shows that predictor models have stable and reliable performance according to all the evaluation metrics. This uniformity shows that the model is not discriminating any class and making the classification consistent and handling the data set well in spite of the imbalances if present.

Accuracy means that in nearly four out of five times the model get its prediction accurately and that's why in every case all around good performance. Precision 79.28% means that most of the positive predictions that model made were correct, i.e. lowrates of falsepositives. Likewise, recall at 79.28% indicates that the model has learned to detect almost 4 out of every 5 real positive instances which is a good balance between sensitivity and selection. The F1 score which is seen as a harmonic mean of precision and recall is 79.26%, this confirm this consistency since there is little difference between these metrics. These results suggest that the Logistic Regression model is quite generalizable to new data, and is not too highly-fitted to specific patterns in the training data. While the 79% performance is excellent for most classification problems, the appropriateness of these metrics is dependent on the context of the application - such as it may be necessary to have higher recall in medical diagnosis or higher precision in fraud detection. Nevertheless, the very close value to all the metrics reveals a well balanced and reliable model which is ready to be used for practical purposes.



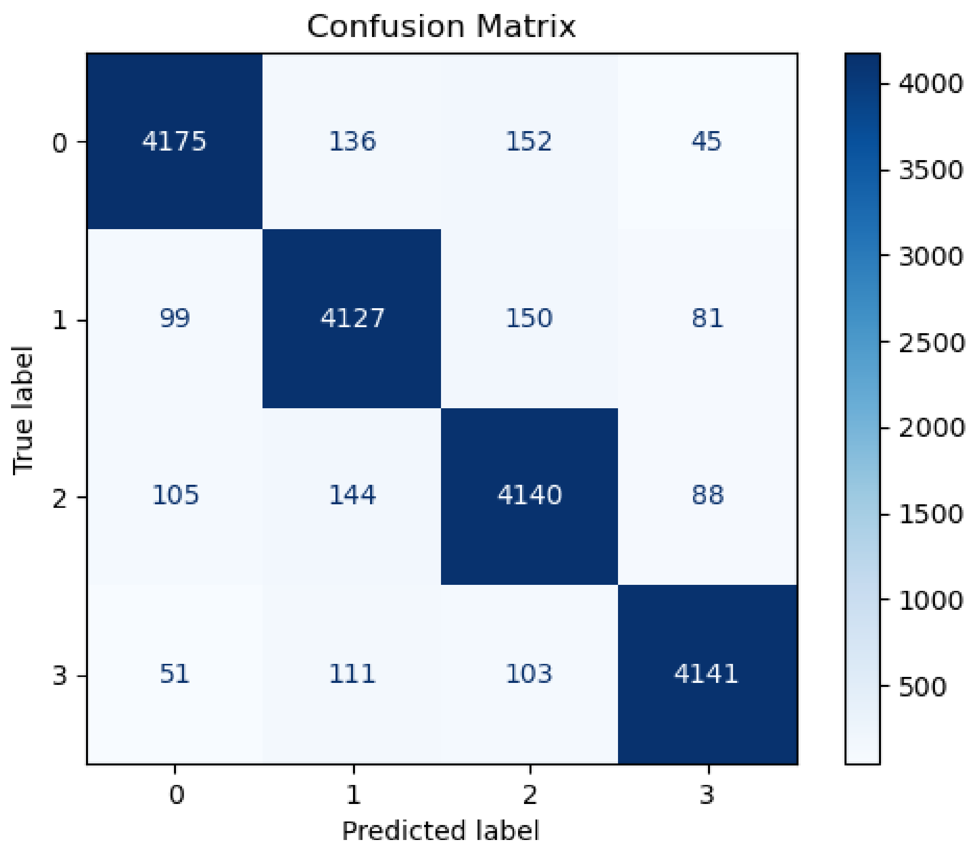
	precision	recall	f1-score	support
0	0.82	0.79	0.80	4508
1	0.77	0.77	0.77	4457
2	0.78	0.77	0.78	4477
3	0.80	0.84	0.82	4406
accuracy			0.79	17848
macro avg	0.79	0.79	0.79	17848
weighted avg	0.79	0.79	0.79	17848

Figure 5.1 Confusion Matrix and Classification Report of LR

Evaluation of Logistic Regression (LR) model on four classes for classification has been further demonstrated by the confusion matrix and classification report. The confusion matrix provides a very detailed breakdown of the amount of instances correctly and incorrectly predicted from each category. While the off-diagonal is referring to cases that are misclassified, the substantial in the main diagonal (3553, 3439, 3457 and 3701) reflects the number of samples that are correctly identified according to the model with respect to each respective class. These results are then distilled in the classification report as well as the usage of certain key performance metrics. For example, precision, for example 0.82 for class 0, where the preseted percentage of predictions of a particular class is positive, describes the reliability of positive predictions for a model. Similarly, recall, i.e. recall--0.79 for class 0, is a measure of the model to detect all actual instances of that class, i.e. in an emphasis on sensitivity. Overall, we can see from the model, it is well balanced and consistent in all classes, and the overall accuracy giving 79% and F1-scores similar in all classes which shows that the model is well distinguishing between the four classes with no much bias towards any particular class.

Random Forest Classifier

With accuracy, recall and f1-score of 92.91%, 92.93%, and 92.91 considering the order, Random forest algorithm (RF) proved to be exceptionally good with an accurate and reliable classifier. For applications such as healthcare, fraud or security, where to get it wrong might be an expensive mistake, these measures indicate that the model is a fair balance between being able to make the right predictions and to produce as few false positives/false negatives as possible. Nearly majority of the true positive are of right. which is shown by high recall. The accuracy and recall do not get affected by the F1-score. There is no skewness on the dataset that needs to be corrected because the model is dealing with all classes properly as reflected by the consistent number from all the metrics. Such consistent performance reveals that the RF model is reliable, generalizes fairly well, and can be applied with some certainty in real world situations which need little post processing and human involvement. The power of ensemble learning in capturing complex and non-linear relationships and mitigating overfitting can be seen in this performance, likely due to careful preprocessing, feature selection and hyperparameter tuning. All things considered, the RF model is a very effective, dependable, and stable method in case used for multi-class classification tasks.



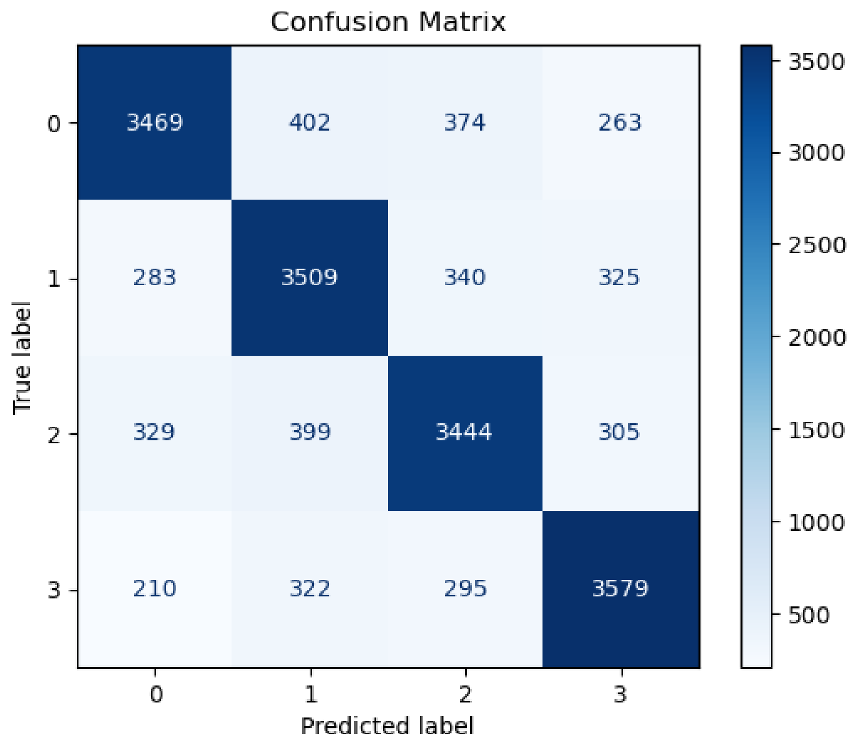
	precision	recall	f1-score	support
0	0.94	0.93	0.93	4508
1	0.91	0.93	0.92	4457
2	0.91	0.92	0.92	4477
3	0.95	0.94	0.95	4406
accuracy			0.93	17848
macro avg	0.93	0.93	0.93	17848
weighted avg	0.93	0.93	0.93	17848

Figure 5.2 Confusion Matrix and Classification Report of RF

The classification report for four class issue and the confusion matrix are used to validate the above-mentioned finding. Small off-diagonal entries represent minimal misclassification while big diagonal (4175, 4127, 4140, 4141) represent a high frequency of right classification. High class-wise accuracy, recall and F1-scores have also been verified by the report, for example, Class 0 has an accuracy of 94%, and a recall of 93%. The Random Forest model is efficient and very robust model on this data with the overall accuracy level of 93% with consistently high macro and weighted average in all metrics.

Decision Tree Classifier

The Decision Tree (DT) model gave us accuracy values, precision value, recall values & f1 score values ranging between 78 to 78.5% which represented moderate performance values of the model. The model operates recursively by dividing up the dataset based on the most informative features to form a tree-like structure where the branches are decision rules and the leaves are class outcomes. The consistency in all the metrics indicates that the predictions are fairly balanced between classes where no class is favoured over the other class. However, the 78% accuracy is also reflective of the fact that a significant percentage of instances are misclassified, which in all likelihood is related to common drawback of a standalone decision tree such as overfitting, sensitivity to the depth of the tree or poor choice of parameters of splitting the nodes. While DTs may be somewhat challenged in highly non-linear or complex feature-target relationships in comparison to the ensemble methods such as Random Forest algorithm, DTs are useful in the application where interpretability is priorities. With fine-tuning of hyperparameters (e.g maximum depth or minimum samples per split or minimum criterion) improvements in the performance can be achieved or ensemble techniques can be used for improved generalisation and reduced variance. Overall, the DT model provides stable and reasonably effectiveness classification, particularly on the issue of transparency.



	precision	recall	f1-score	support
0	0.81	0.77	0.79	4508
1	0.76	0.79	0.77	4457
2	0.77	0.77	0.77	4477
3	0.80	0.81	0.81	4406
accuracy			0.78	17848
macro avg	0.78	0.78	0.78	17848
weighted avg	0.78	0.78	0.78	17848

Figure 5.3 Confusion Matrix and Classification Report of DT

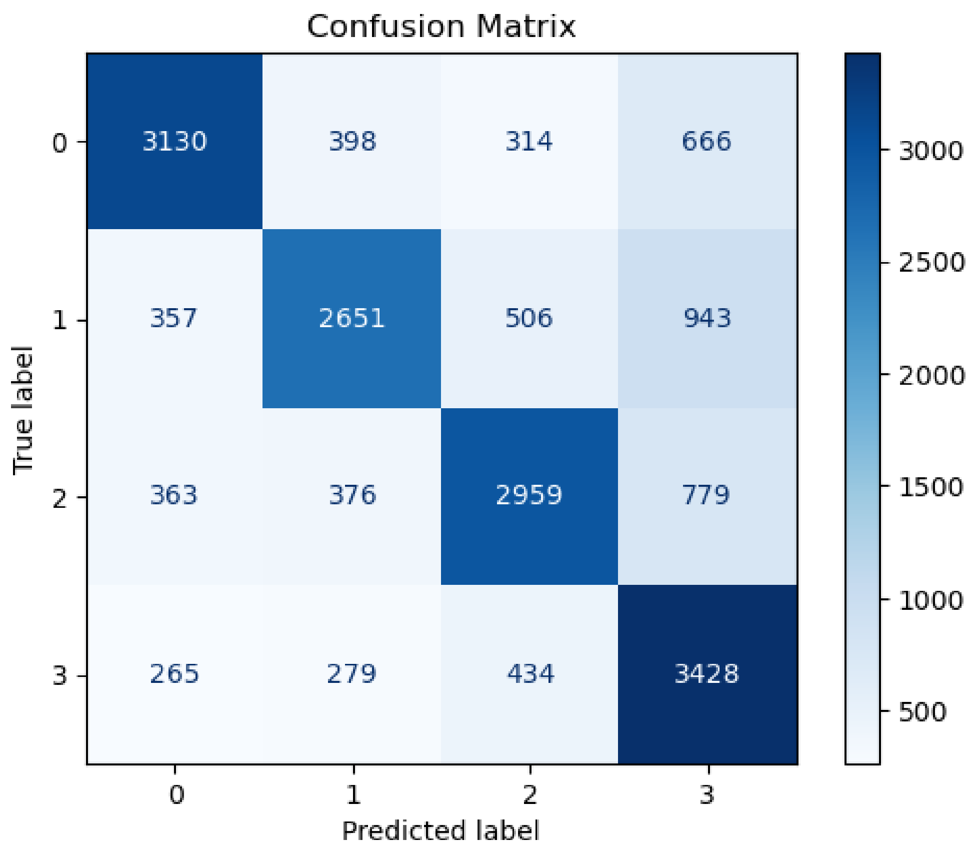
The confusion matrix and the classification report bear these observations out in the four class problem. Diagonal values (3469, 3509, 3444, 3579) represent majority of correct prediction and off diagonal misclassifications are pretty low. Class-wise precision, recall and F1-scores are in between 0.77 to 0.81, which implies that the performance is balanced across all classes. The overall accuracy of 78% confirms that the DT model is a model that can be used consistently as an interpretable machine learning technique, though additional tuning or ensemble methods could be used to improve the model's predictive capability even further.

XGB Classifier

The extreme gradient boosting algorithm (XGB) model has the highest result of 68.17% as accuracy, 69.25% as precision, 68.17% as recall, and F1-score as 68.19%. The classification performance results have been evaluated as moderate. XGB is a method of gradient boosting and is generated decision trees through each other in which the each new tree corrects the errors of the previously generated tree. The results show that the model is able to predict correctly about two-thirds of the instances, but there is a lot of room for improvement, which could be due to class imbalance, overlapping features, noisy data or not discriminative enough.

The fact that the recall is marginally larger than the precision lets us know that the model is in the habit of being conservative in making positive predictions, with fewer false positive predictions, which would be good in situations where their false predictions are costly, such as in medical diagnosis or fraud detection. However, it appears from the recall that the model is not able to capture a significant portion of the actual positives and thus, there may be a need to make a few adjustments to be able to capture all of the relevant cases. The F1-score, as it is near both the accuracy and the recall, has a good trade off between both of these measures.

On the whole the XGB classifier is quite reliable, more so in the negative than in the positive way (the fact is that this way the classifier takes a bit more ifines). Improving performance could be done by hyperparameter search (i.e. learning rate, tree depth, no of estimators), feature engineering or balancing the class using SMote or by cross validation for better generalization.



	precision	recall	f1-score	support
0	0.76	0.69	0.73	4508
1	0.72	0.59	0.65	4457
2	0.70	0.66	0.68	4477
3	0.59	0.78	0.67	4406
accuracy			0.68	17848
macro avg	0.69	0.68	0.68	17848
weighted avg	0.69	0.68	0.68	17848

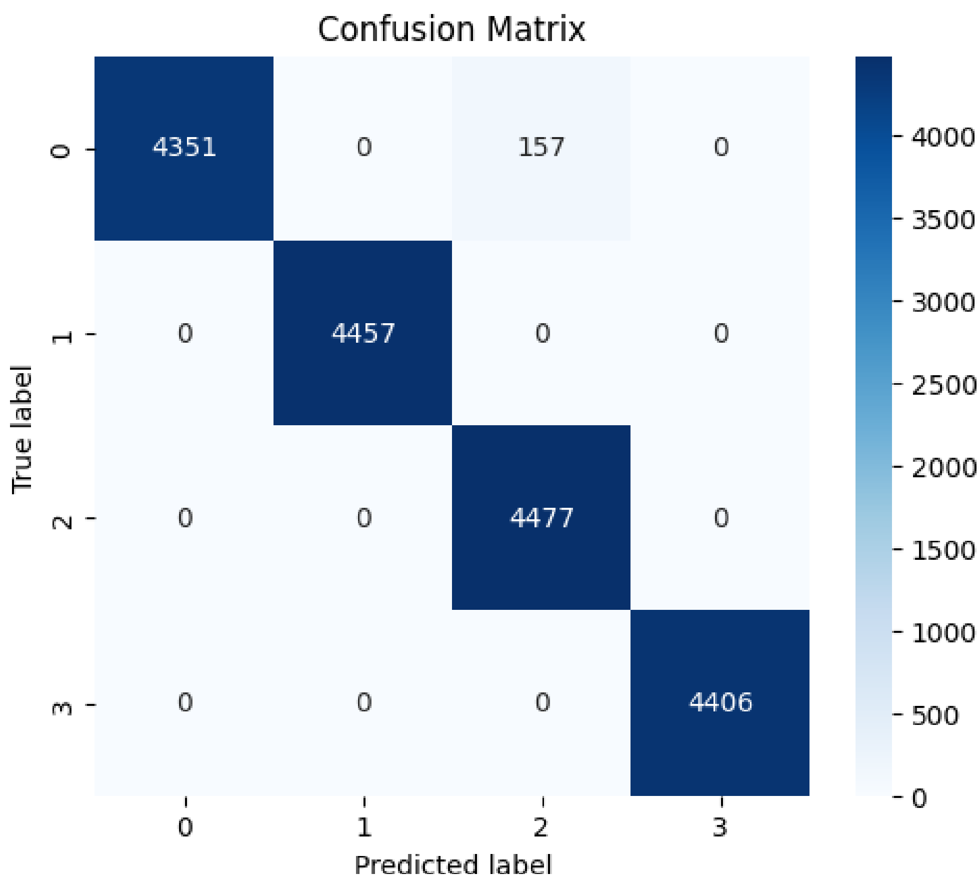
Figure 5.4 Confusion Matrix and Classification Report of XGB

The confusion matrix and the classification report gives us a bit more information. Large off-diagonal figures show that there are lots of misclassifications, e.g., of class 1 classified as class 3 (943 times) and of class 0 classified as class 3 (666 times). While the model has a high recall for class 3 (0.78), the recall has a low precision (0.59) with inconsistent performance across class. These results suggest that the XGB model can be further optimized if a better class separation and overall model dependability were to be guaranteed.

Extra Trees Classifier

The Extra Trees Classifier (ETC) was found to give some excellent results with a 99.12% accuracy, 99.15% precision, 99.12% recall and F1-score 99.12%, which shows very close to perfect performance in all the evaluation metrics. These results show that the model has been able to capture the underlying patterns in the dataset and that it generalizes to unseen data. The high precision and recall can be assumed to have minimal false positive and false negative and the balanced F1-score means that all the classes are equally performed without any favouritism to one class over the other.

ETC is an ensemble tree-based technique akin to Random Forest with an additional randomness for learning, which takes into consideration best splits for all features instead of random set of features, which tends to reduce variance and makes it more robust. The powerful performance in this case is the result from the quality of the input feature and proper preprocessing, for example, missing value, feature encoding, scaling. The fact that accuracy and recall are close to each other suggests that the data set is fairly balanced or the model is a good way of dealing with class imbalance.



	precision	recall	f1-score	support
0	1.00	0.97	0.98	4508
1	1.00	1.00	1.00	4457
2	0.97	1.00	0.98	4477
3	1.00	1.00	1.00	4406
accuracy			0.99	17848
macro avg	0.99	0.99	0.99	17848
weighted avg	0.99	0.99	0.99	17848

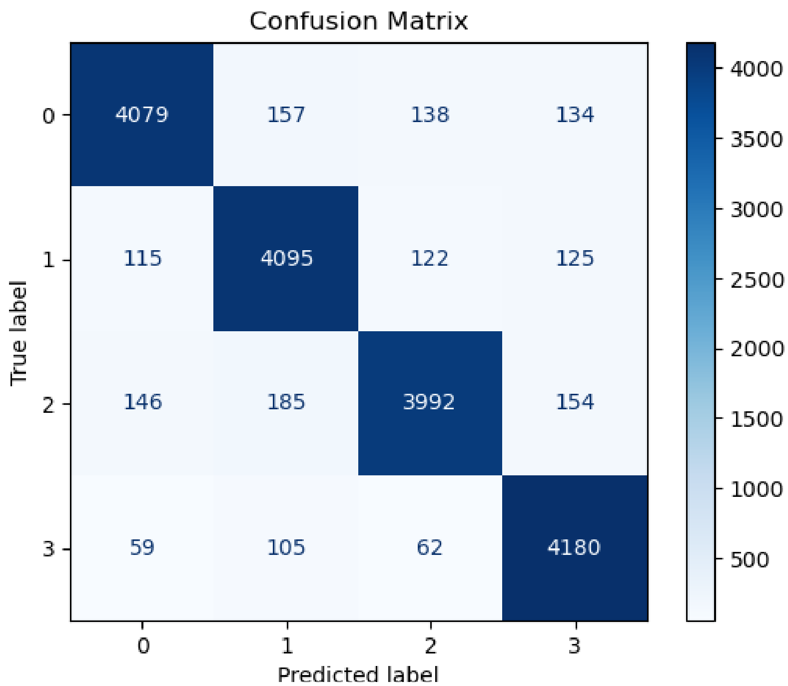
Figure 5.5 Confusion Matrix and Classification Report of ETC

Such high performance makes ETC ideal for real-world applications in which both false positives and negative come at high costs, like fraud detection, quality control or medical diagnostics. To make the reliability even more sure, some holdout set or k fold cross validation is advised to make certain that the model is actually capturing a few of the actual patterns and not overfitting. The near-perfect metrics of the ETC demonstrate the efficiency of ensemble tree concepts in complex classification problems with a well prepared data and optimized hyperparameters that can provide a well balanced and reliable predictive strategy.

MLP Classifier

Results With an accuracy, precision, recall and F1-score of 91.58%, 91.61%, 91.58% and 91.57%, respectively, Multi-Layer Perceptron (MLP) Classifier performed well. The model learns patterns in the data without becoming biased toward any kind of the class, as well observed by the near alignment of these intrapanel values, which exhibit stability and reliability in all of the assessment measures. The low gap between precision and recall indicate that the MLP, a type of feedforward neural network, would be able to balance a situation between false positive and false negative by using the characteristics of the non-linear activation functions of the hidden layers to represent complex relationships between the data.

Good hyperparameter tweaking, the use of regularization techniques such as dropout or L2 penalties, and perhaps a big enough dataset of decent samples for the model to generalize effectively all the above are reflected into such performance. Of the options, 92 out of 100 choices of the images are right and the accuracy is around 91.6%. The model isn't sacrificing one metric for another as seen from its accuracy, recall & F1-score. In industries such as healthcare, banking or risk assessment, where mistakes in one sector could have major consequences, it is a very valuable balance.



	precision	recall	f1-score	support
0	0.93	0.90	0.92	4508
1	0.90	0.92	0.91	4457
2	0.93	0.89	0.91	4477
3	0.91	0.95	0.93	4406
accuracy			0.92	17848
macro avg	0.92	0.92	0.92	17848
weighted avg	0.92	0.92	0.92	17848

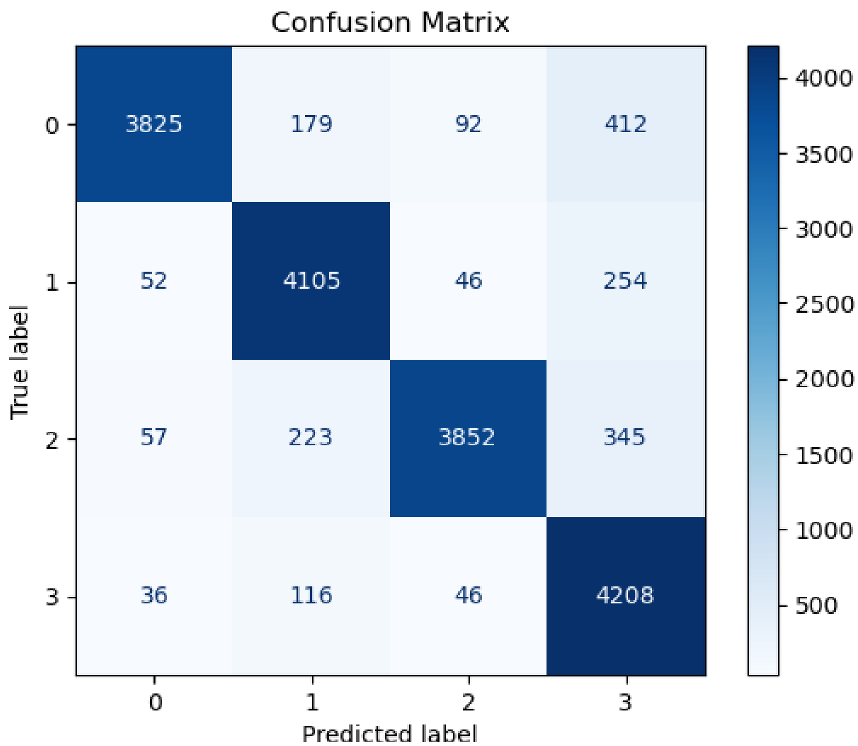
Figure 5.6 Confusion Matrix and Classification Report of MLP

Consistency in finding is also an indication of the use of stabilizing techniques (batch normalization) and proper preparation (feature selection, normalization, weight initialization). The design of the MLP appears to be at the correct depth in order to prevent over fitting and under fitting and learn significant patterns. All things taken into consideration, the consistently high result is a robust and well-trained model that is good in generalizing to new data that is, in turn, applicable for practical use in a real world setting.

K Neighbors Classifier

With accuracy being 89.58%, precision as 90.35%, recall as 89.58%, and F1-score as 89.65%, a good performance of K-Nearest Neighbors (KNN) classifier was obtained on the dataset. These results show the dependability and efficacy of this model under a range of assessment parameters. Nearly nine out of ten predictions were correct, and the high accuracy suggests that most of the positive predictions were correct, reducing the rate of false positives, an important measure in areas such as risk assessment and fraud detection as well as healthcare. The recall demonstrates that nearly all the true positive examples were correctly captured by the model, which implies that there were few relevant cases missed by the model. The F1-score that is very similar to the accuracy and recall shows a performance that is neutral and impartial.

The K value, suitable distance parameter, and so preprocessing such as feature scale are all factors to the efficacy of the model. The findings also suggest that potential issues such as class imbalance or noise are well addressed, and that the dataset has definite class boundaries which may be utilised by KNN. All things considered, the KNN classifier had a precise, reliable, and easy to comprehend predictions to make it a sensible option for this classification challenge.



	precision	recall	f1-score	support
0	0.96	0.85	0.90	4508
1	0.89	0.92	0.90	4457
2	0.95	0.86	0.90	4477
3	0.81	0.96	0.87	4406
accuracy			0.90	17848
macro avg	0.90	0.90	0.90	17848
weighted avg	0.90	0.90	0.90	17848

Figure 5.7 Confusion Matrix and Classification Report of KNN

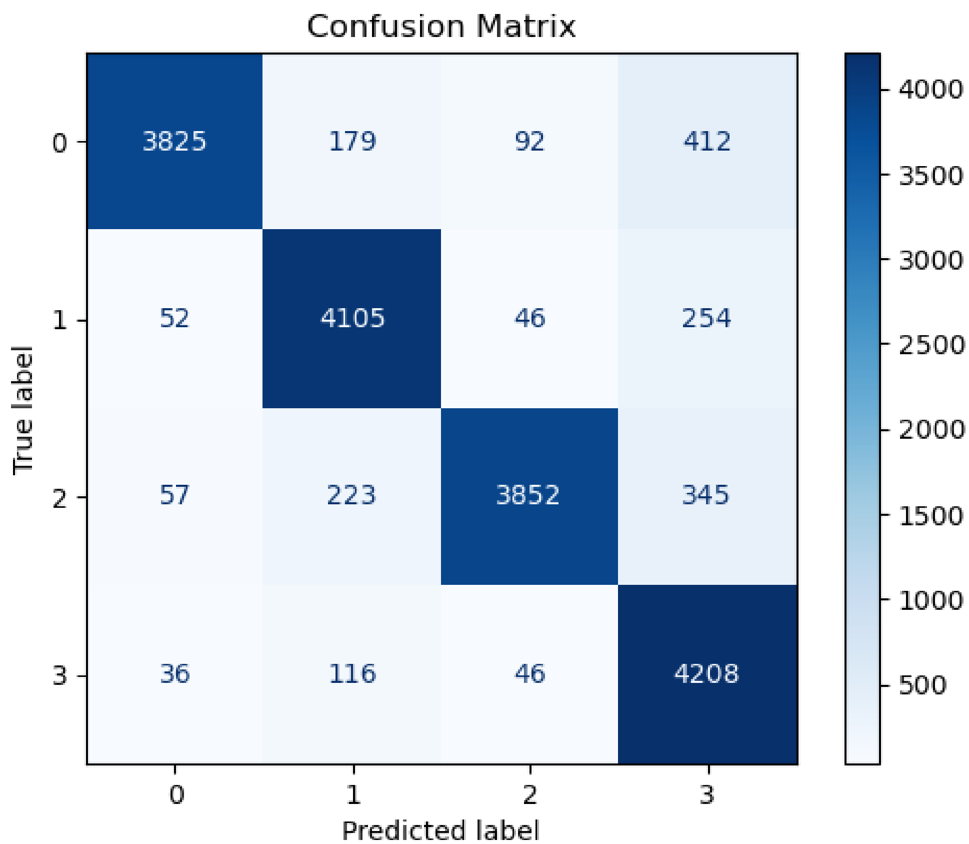
The categorization report and confusion matrix are additional proof of the effectiveness of the model. The majority of cases were correctly identified (as can be shown by the large values along the main diagonal (3825, 4105, 3852, 4208)). There were various misclassifications which were especially high in class 3 which had a high recall (0.96) but a poor accuracy (0.81) which indicated that it was often mis-classified as other classes. Classes 0 and 2, on the other hand, proved reliable forecasts with high accuracy (0.96 and

0.95). The overall indicators like weighted F1-scores of 0.90 and balanced macro show the strong and extensive consistencies of the model for all classes, but with some class- specific trade-offs.

LGBM Classifier

The Light Gradient Boosting Machine (LGBM) proved to be a powerful model on the data set with the accuracy of 89.89%, the precision of 89.94%, the recall of 89.89%, and the F1-score of 89.9%. These metrics show that the model reasonably captures complex patterns and has a balanced prediction of each class. Accuracy gives us the satisfaction that almost 90% of all cases were accurately classified while a high precision indicates that most positive predictions were accurate which reduces false positives - very important in areas such as medical diagnosis, detecting fraud, or quality control. The recall value tells us that the model has managed to recognize nearly all the true positive cases which reduce the number of false negatives which is essential in high-stakes applications. The F1-score, which is very close to accuracy and recall, indicates an equilibrium performance without bias to any given performance.

LGBM's good performance can be attributed to the gradient boosting mechanism that can correct the errors of previous trees one by one, and the optimized tree growth with leaf-wise, histogram-based splitting, and efficient handling of categorical features. These factors help to reduce the overfitting and make the model can generalize well. The similarity between the metrics also indicates good preprocessing, scaling of features and tuning of hyperparameters such as learning rate, number of leaves and boosting rounds.



	precision	recall	f1-score	support
0	0.96	0.85	0.90	4508
1	0.89	0.92	0.90	4457
2	0.95	0.86	0.90	4477
3	0.81	0.96	0.87	4406
accuracy			0.90	17848
macro avg	0.90	0.90	0.90	17848
weighted avg	0.90	0.90	0.90	17848

Figure 5.8 Confusion Matrix and Classification Report of LGBM

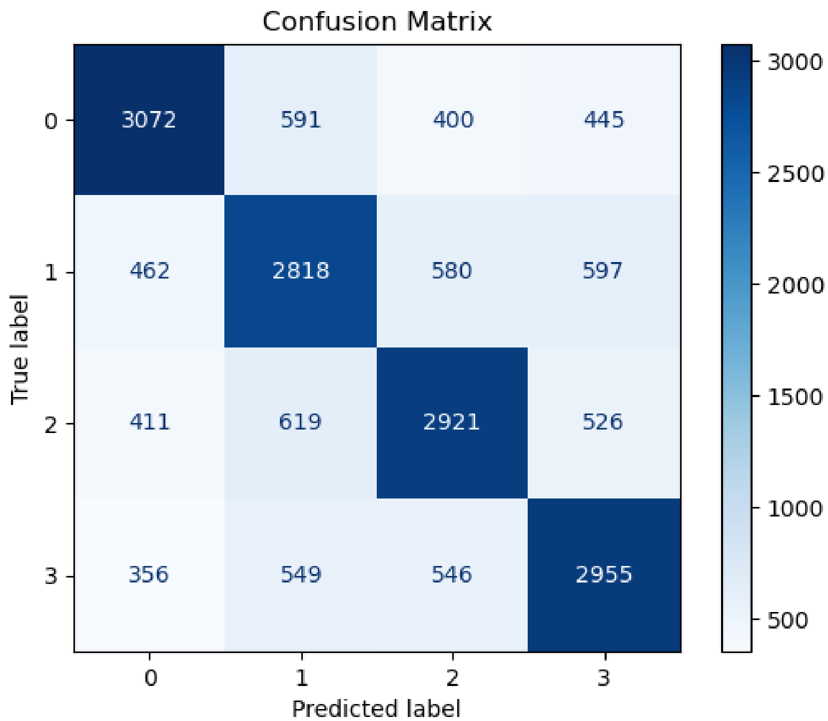
The LGBM classifier reveals to be a robust and trustworthy model that is capable of providing exact, balanced and high-performing predictions. Its capability of processing complex interactions and stability makes it applicable so that it can be used for tasks in the real world that require reliable classification.

Confusion matrix and classification report are formed to verify the performance of the model. Large diagonal values indicate that most were correctly classified and minor misclassifications between certain classes indicate areas for improvement. Class specific metric indicates that some classes perform a bit better recall at the cost of precision but the overall macro and weighted F1-scores result to a reliable and consistent performance across all categories, namely 0.90.

CatBoost Classifier

CatBoost classifiers, an algorithm making some adjustments specific to categorically defined features and preventing overfitting, performed admirable on the data set. The accuracy, precision, and recall scores of them were 65.92%, 66.01%, 65.92%, and 65.95%, respectively, and their F1 score was 65.95%. These metrics indicate that the model is in the range of correct prediction in around two-thirds cases but the model is not in a position to effectively yield complex patterns when compared to more powerful models like LGBM or Extra Trees. Since the accuracy alone is not enough to consider class imbalance and type of errors, it is better to evaluate the model performance by measuring precision, recall, and F1-score which give a more overall picture of a model's performance.

According to accurate, about 66% of positive predictions were correct, demonstrating that while the model does not have large false positive, it indeed cannot be very reliable based on the consequences of being positive. Recall means that the model is catching two out of three of the true positives in sensitive applications, such as fraud detection or medical diagnostics, implies that you were losing out on two out of three actual positives. In terms of accuracy, recall, F1-score of the overall performance is a balanced but normal performance.



	precision	recall	f1-score	support
0	0.71	0.68	0.70	4508
1	0.62	0.63	0.62	4457
2	0.66	0.65	0.65	4477
3	0.65	0.67	0.66	4406
accuracy			0.66	17848
macro avg	0.66	0.66	0.66	17848
weighted avg	0.66	0.66	0.66	17848

Figure 5.9 Confusion Matrix and Classification Report of CatBoost

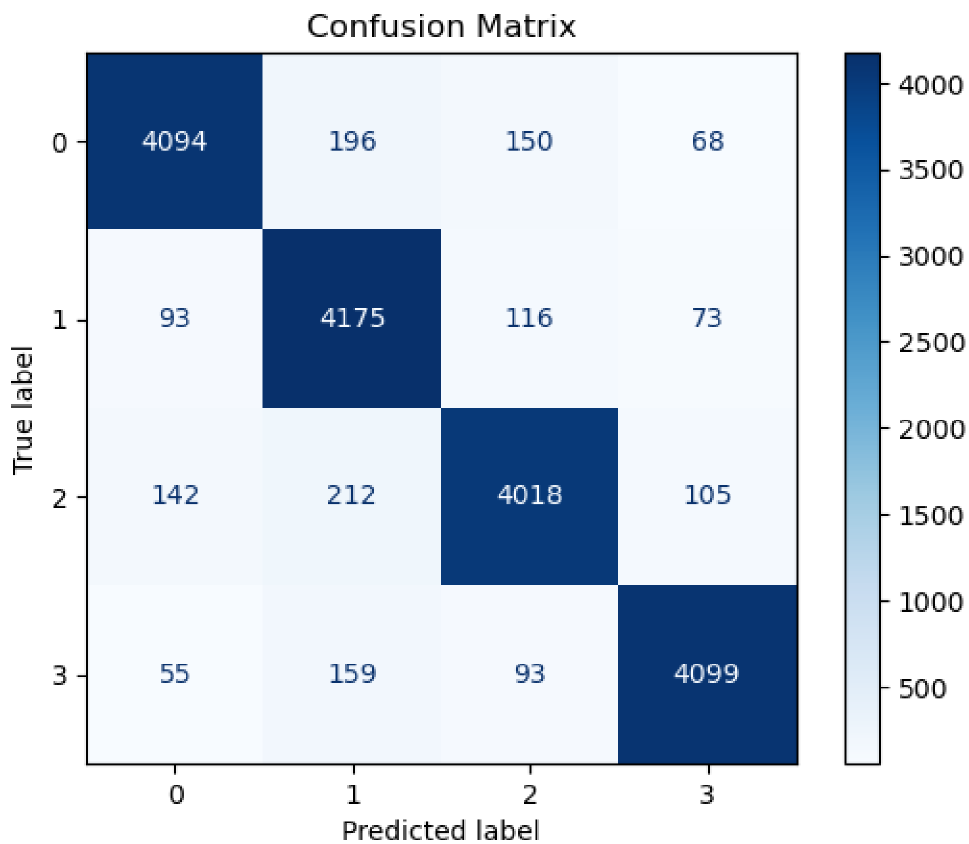
Even though CatBoost can manage categorical variables, the bad performance may be created based on the lack of training data, noise, imbalanced classes, complicated feature interactions that cannot be captured by the model, or constrained hyperparameter tuning. These findings mean that even with its potential, the CatBoost may need more success with more feature selection, preprocessing and hyperparameter optimization.

The results from the confusion matrix and classification report support these results. While off-diagonal misclassfections, such as Class 0 being more often called Class 1 (591) or 3 (445), indicate the classifiers weak ability to discriminate, the important diagonal values (3072 2818 2921 and 2955) indicate a very respectable frequency of correct predictions. Class-wise measures range between 0.62 and 0.71 which should tell us that even though the model's predictions are well balanced, they often are not as good as the ones of the best classifiers.

Artificial Neural Network (ANN)

Inspired by the working of the human brain, the Artificial Neural Network (ANN) is a powerful tool for determining complex nonlinear correlations in the data. It demonstrated good and high predictive results perfectly on the studied data set, with an accuracy, precision, recall and F1-score of 91.81%. This shows the generalisation ability of the model to some extent derived from the training data to the test data as the model adhered to the principle of recognising in over nine out of 10 cases correctly. Precision, recall and F1-score provide information about the error types as well as class specific behavior, whereas accuracy provides an elemental performance synopsis.

With a precision of 91.81%, the model demonstrated its reliability in reducing false positives, which is important in sensitive applications such as fraud detection or medical diagnostics. Almost all the positive predictions turned out to be correct. In the case of instances that are missing, but that could potentially have large consequences, a model's recollection 91,81% is emphasized and demonstrates the model's capability to find almost all true positives without increasing the false negatives effect. Numbers for accuracy, precision and recall are nearly equivalent which show that ANN is impartial and well-balanced.



Class 0	0.93	0.91	0.92	4508
Class 1	0.88	0.94	0.91	4457
Class 2	0.92	0.90	0.91	4477
Class 3	0.94	0.93	0.94	4406
accuracy			0.92	17848
macro avg	0.92	0.92	0.92	17848
weighted avg	0.92	0.92	0.92	17848

Figure 5.10 Confusion Matrix and Classification Report of ANN

Being a balance between accuracy and recall, F1-score (also 91.81%) makes the performance of the model reliable. This balance is only possible to achieve stability and quality predictions without overfitting and is the result of a good design architecture and proper regularization and hyperparameter tuning.

All in all, the ANN performance is better than the simpler model, such as KNN or Catboost, and ANN model can communicate with complex features efficiently and also give the accurate and quality prediction. This is confirmed by the confusion matrix which shows that most predictions made are on the main diagonal (4094, 4175, 4018, 4099). The classification report further depicts good performance in all the classes with good class-wise accuracy, recall and F1 score of more than 0.88.

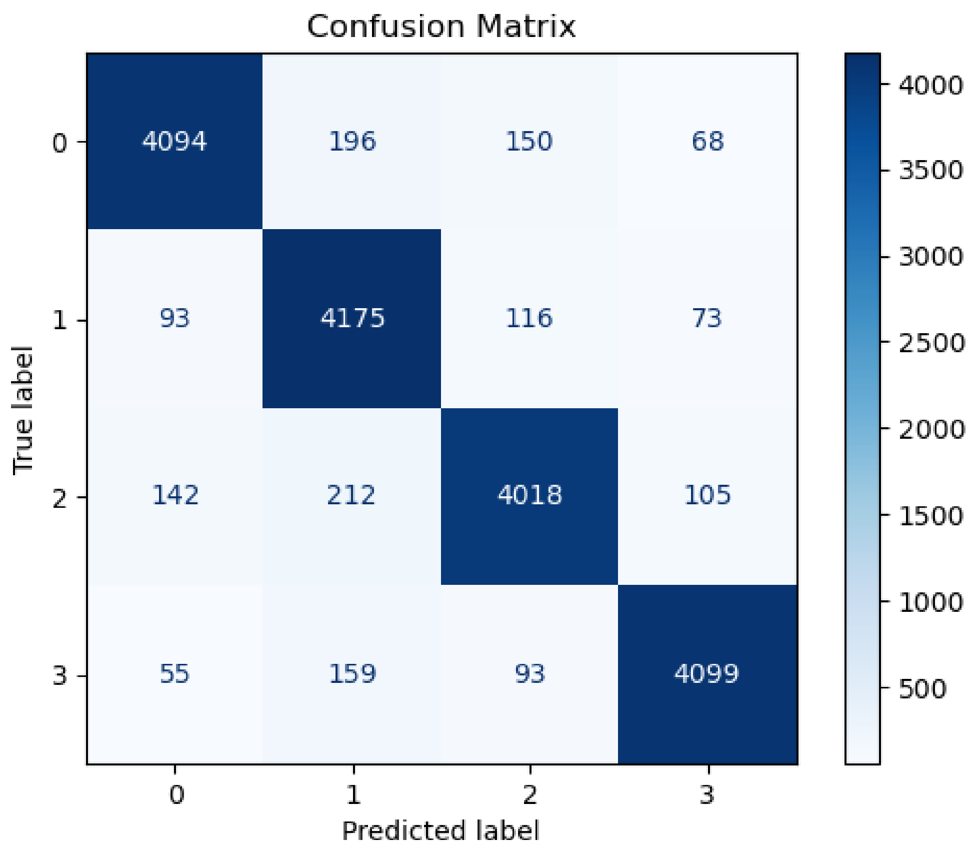
Convolutional Neural Network (CNN)

A special kind of neural network is convolutional neural networks (CNNs) which are specialized in the task of extracting features, extremely good at processing sequential and organized input such as grids or pictures. It showed excellent and balanced performance in the evaluated data set having accuracy, precision, recall and F1-score of 92%. This means that the model is correct 92 times out of 100.

With the accuracy of 92%, the network is reliable in reducing the false positives which makes it suitable for applications like fraud detection or medical diagnoses where false positive predictions are costly. The convolutional layers, filters, pooling and non-linear activations of the CNN system make the CNN system more reliable in its predictions by identifying minute patterns and decreasing the noise.

With a recall of 92% the model at least is able to reduce false negative by capturing almost all the real positive cases and. In situations where the effects of missing some number of instances can be disastrous, high recall is imperative. The ability of a network for dealing in a unbiased way in both positive or negative situations can be demonstrated by the balanced performance for the parameters such as accuracy, precision and recall.

The general consistency of developed model is validated in 92% obtained F1-score that provides balance between recall and accuracy in one and coherent index of efficacy. This high performance is retained and overfitting is prevented, by use of regularization techniques, dropout techniques, batch normalization and where needed, data augmentation.



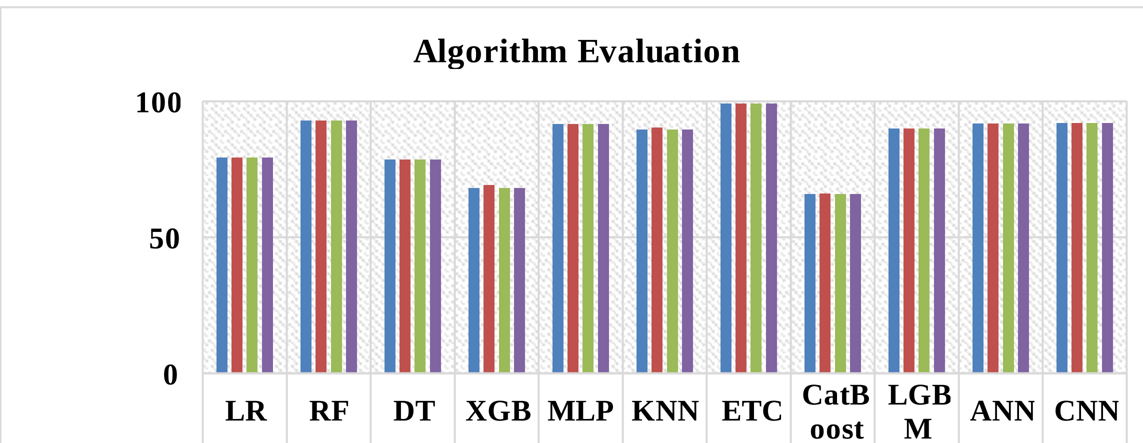
	precision	recall	f1-score	support
Class 0	0.93	0.91	0.92	4508
Class 1	0.88	0.94	0.91	4457
Class 2	0.92	0.90	0.91	4477
Class 3	0.94	0.93	0.94	4406
accuracy			0.92	17848
macro avg	0.92	0.92	0.92	17848
weighted avg	0.92	0.92	0.92	17848

Figure 5.11 Confusion Matrix and Classification Report of CNN

All things considered, the CNN has a great feature extraction, is little complex association and has dependable generalization to new data. It works a little better, and more consistently, than more straightforward classifiers, such as KNN or CatBoost that are because it is able to pick up finer patterns. This is then supported with the confusion matrix which displays high values along the main diagonal (4094, 4175, 4018, 4099). Additionally, it can be found in the classification report that the precision, recall and F1 scores are balanced across all classes (0.88-0.94) which shows the suitability of CNN for tasks that require accurate and reliable predictions.

4.2 Discussion

According to the mentioned measures of accuracy, precision, recall, and F1-score, a comparative analysis of machine learning and deep learning algorithms can be an important source of information regarding the relative strengths and weaknesses of both models to process the dataset under investigation. The evaluated ten algorithms selected included Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), XGBoost (XGB), Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), Extra Trees Classifier (ETC), CatBoost, Light Gradient Boosting Machine (LGBM), Artificial Neural Network (ANN), and Convolutional Neural Network (CNN) and performed differently because of the probable complexity of the provided dataset consisting of nonlinear relationships, interactions between the features, and differences in These results show the importance of selecting the algorithm, optimizing the hyperparameter, and the model suitability to the characteristics of the input data to achieve the optimal predictive performance.



	LR	RF	DT	XGB	MLP	KNN	ETC	CatBoost	LGBM	ANN	CNN
Accuracy	79.28	92.91	78.44	68.17	91.58	89.58	99.12	65.92	89.89	91.81	92
Precision	79.28	92.93	78.49	69.25	91.61	90.35	99.15	66.01	89.94	91.81	92
Recall	79.28	92.91	78.44	68.17	91.58	89.58	99.12	65.92	89.89	91.81	92
F-1 Score	79.26	92.91	78.44	68.19	91.57	89.65	99.12	65.95	89.9	91.81	92

Figure 5.12 Evaluation Comparision

The models were compared and it was revealed that deep learning, ensemble techniques and the classical machine learning all had varying performance patterns. The mid-range baselines of the Decision Tree (78.44 percent) and the Logistic Regression (79.28 percent) were limited by overfitting and linear premises, respectively. Ensemble algorithms with bagging, unpredictability, and robust decision boundaries to learn complex interactions between features performed better than simple models, specifically the Random Forest (92.91) and Extra Trees Classifier (99.12). Due to unfavorable hyperparameters or dataset characteristics, XGBoost (68.17) and CatBoost (65.92) as gradient boosting algorithms did not perform well, but LightGBM achieved good accuracy (89.89) under a successful computing. The instance-based KNN that utilized local neighborhood structures had a good performance (89.58%). The other deep learning models did not perform as well as CNN (92%), as its performance was largely due to its efficient feature extraction, and ANN (91.81%), MLP (91.58%), and CNN (92%) showed strong performance also, due to their ability to provide hierarchical patterns and nonlinear interactions. The neural network and ensemble trees performed optimally in modeling complex datasets and constantly scored highest in accuracy, balanced precision, recall, and F1-scores. Frugal baselines were traditional and some boosting models.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

The development of a powerful tool of sentiment analysis has significant social implications, particularly, in the context of enhanced decision-making and better understanding of the opinion of the population. By carefully scrutinizing the posts of social media, customer reviews and other user-generated content, businesses, legislators and social organizations can better meet, satisfy and address the needs, wants, and issues of the society. Better sentiment detection would potentially be beneficial in monitoring mental health trends, identifying new issues affecting society and tailoring services to meet the needs of a community. Moreover, technology also enables one to react to disasters, political campaigns, and other situations affecting the health of people more promptly and more wisely which ultimately makes the society more united, sensitive, and aware.

5.2 Impact on Environment

Sentiment analysis indirectly affects the environment even though it primarily operates with digital data and indirectly impacts the computer resources. Complex machine learning and deep learning models, particularly deep neural networks, require significant energy to be trained and when not optimized may be carbon-emitting when running on a large-scale cluster of GPUs. Nevertheless, energy consumption can be reduced by streamlining training pipelines, training on existing models, and implementing efficient algorithms. The understanding obtained through sentiment analysis can also be applied to environmental conservation, and the data obtained can be used to promote environmental awareness, follow the way people perceive sustainability efforts, and guide legislation that promotes sustainable behavior.

5.3 Ethical Aspects

The ethical concerns in this study focus on justice, privacy and relevant use of data. All the information contributed by users, such as reviews and comments on a social network, need to be anonymized to retain personal identities and comply with the legislation on data protection, such as the GDPR. The research does not practice prejudice or discrimination because of employing class balancing techniques, ensuring that minority classes are represented properly, and ensuring that the model projections do not contribute to the development of adverse views. Transparency in model selection, preprocessing and assessment also ensure accountability. More so, ethical use must be monitored closely to prevent misuse like influencing people or infringing the privacy of an individual.

5.4 Sustainability Plan

The sustainable strategy focuses on long-term usage and creation of energy saving models. The study lowers energy consumption and carbon footprint through optimization of computing resources through pre-trained models, batch processing and efficient algorithms such as LightGBM or ETC. To ensure sustainability of the resources, frequent model renewal, which employs incremental learning, can reduce the need to train afresh. Also, the flexibility of the framework allows companies to apply it to diverse industries without the necessity of major re-engineering, which increases its social value without reducing operational and ecological sustainability.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

In order to attain precision and trustworthy sentiment classification, the analysis offered in this study is a detailed sentiment analysis architecture that incorporates deep learning networks, ensemble approach, and conventional machine learning algorithms. Various models were systematically tested on a processed dataset of Twitter which was TF-IDF vectorized and balanced using SMOTE including Logistic Regression, Decision trees, KNN, random forest, Extra Trees Classifier, gradient boosting algorithms and neural networks like ANN, MLP and CNN. The results of comparative analysis showed that the deep learning models and ensemble tree models (CNN, ANN, MLP, ETC, RF) have never recorded low accuracy, preciseness, recall and F1-score than the simple models. CNN and ETF were the most secure ones. To reveal the complex patterns and nonlinear interactions, the study refers to the significance of preprocessing and feature engineering and a fine selection of models. It renders it, altogether and endlessly, a repeatable and adaptable pipeline that provides a tradeoff between predictive execution and computational performance in giving it worthwhile application to the field of social media monitoring, customer feedback analysis and decision making processes in all the different fields.

6.2 Conclusions

To have a good and accurate sentiment categorization, the current study illustrates the functionality of an overall sentiment analysis pipeline that integrates both conventional machine learning techniques and the emerging deep learning techniques. Though extensive preprocessing, feature engineering and feature balancing provided a high-quality input during model development, a comparative analysis showed that neural networks (MLP), and ensemble models (Random Forest, Extra Trees, gradient boosting) often performed better than simpler models. The results show that the processing resources and the complexity of the dataset should be taken into account in the selection of the method, and the deep learning can detect weak patterns, and conventional models are fast to understand. Overall, the pipeline offers a generic template to be applied in a large variety of applications, including market sentiment analysis, social media analytics, and customer feedback analysis, which demonstrates that both a diligent data preparation process and a diligent model selection should be undertaken in order to achieve useful, reliable data.

6.3 Implication for Further Study

Future studies in this area have to aim at areas that are critical to improve the limitations and expand scope of analysis:

- **Multimodal Data integration:** To get a more close insight about the topic, integrate text and visual data and other information.
- **Different Datasets:** To minimize bias, as well as to generalize the findings more widely, sample culturally and linguistically distinct datasets.
- **Longitudinal Studies:** In order to explore the dynamics of change and establish more robust cause and effect relationships, perform a long-term study.
- **Real-World Applications:** To further the theory-practice gap, consider a case of applying the research findings to other areas of work, including healthcare, business, or education.
- **Ethical Does**: When establishing the models, take into account ethics, openness and justice in particular when processing sensitive information.
- **Hybrid Approaches:** Get more productive and flexible: Hybrid approaches are improved, e.g. by more traditional methods with federated or transfer learning.
- **Cross-Sector Collaboration:** To make sure that research can solve sensible issues, collaborate with practitioners, legislators, and business executives.
- **Reproducibility:** To guarantee stability and similarity of results of a study whenever repeating a study, formulate standard structures and standards.
- **Societal Implication:** Be actively engaged in addressing the social effects of technology in general, such as privacy and inclusion.

Reference:

[1] Luo, X. (2021). Efficient English text classification using selected machine learning techniques. Alexandria Engineering Journal, 60(3), 3401–3409. https://doi.org/10.1016/j.aej.2021.02.009

[2] [2] Ara, J., Hasan, M. T., Al Omar, A., & Bhuiyan, H. (2020, June 1). Understanding customer sentiment: Lexical analysis of restaurant reviews. IEEE. https://ieeexplore.ieee.org/abstract/document/9230712

[3] Alharbi, L. M., & Qamar, A. M. (2022). Arabic sentiment analysis of eateries' reviews using deep learning. Ingénierie des Systèmes d'Information, 27(3), 503–508. https://doi.org/10.18280/isi.270318

[4] Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. Procedia Computer Science, 87, 44–49. https://doi.org/10.1016/j.procs.2016.05.124

[5] Hung, L. P., & Alias, S. (2023). Beyond sentiment analysis: A review of recent trends in text-based sentiment analysis and emotion detection. Journal of Advanced Computational Intelligence and Intelligent Informatics, 27(1), 84–95. https://doi.org/10.20965/jaciii.2023.p0084

[6] Zhou, Q., Xu, Z., & Yen, N. Y. (2019). User sentiment analysis based on social network information and its application in consumer reconstruction intention. Computers in Human Behavior, 100, 177–183. https://doi.org/10.1016/j.chb.2018.07.006

[7] Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. Computer Science Review, 41, Article 100413. https://doi.org/10.1016/j.cosrev.2021.100413

[8] Singh, N. K., Tomar, D. S., & Sangaiah, A. K. (2020). Sentiment analysis: A review and comparative analysis over social media. Journal of Ambient Intelligence and Humanized Computing, 11, 97–117. https://doi.org/10.1007/s12652-018-0862-8

[9] Alhojely, S. (2016). Sentiment analysis and opinion mining: A survey. International Journal of Computer Applications, 150(6), 22–25. https://doi.org/10.5120/ijca2016911545

[10] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135. https://doi.org/10.1561/1500000011

[11] Verma, P., Bhardwaj, T., Bhatia, A., & Mursleen, M. (2023). Sentiment analysis using SVM, KNN and SVM with PCA. In T. Bhardwaj et al. (Eds.), Artificial Intelligence in Cyber Security: Theories and Applications (Vol. 240). Springer. https://doi.org/10.1007/978-3-031-28581-3_5

[12] Park, S., Cho, J., Park, K., & Shin, H. (2021). Customer sentiment analysis with more sensibility. Engineering Applications of Artificial Intelligence, 104, 104356. https://doi.org/10.1016/j.engappai.2021.104356

[13] Gunasekaran, K. P. (2023). Exploring sentiment analysis techniques in natural language processing: A comprehensive review. arXiv. https://doi.org/10.48550/arxiv.2305.14842

[14] Hamad, M. M., Salih, M. A., & Jaleel, R. A. (2021). Sentiment analysis of restaurant reviews in social media using Naïve Bayes. Applications of Modelling and Simulation, 5, 166–172. https://arqiipubl.com/ojs/index.php/AMS_Journal/article/view/304

[15] Cui, J., Wang, Z., Ho, S. B., et al. (2023). Survey on sentiment analysis: Evolution of research methods and topics. Artificial Intelligence Review, 56, 8469–8510. https://doi.org/10.1007/s10462-022-10386-z

[16] Rajasekaran, R., Kanumuri, U., Kumar, M. S., Ramasubbareddy, S., & Ashok, S. (2019). Sentiment analysis of restaurant reviews. In S. Satapathy et al. (Eds.), Smart Intelligent Computing and Applications (Vol. 105). Springer. https://doi.org/10.1007/978-981-13-1927-3_41

[17] Krishna, A., Akhilesh, V., Aich, A., & Hegde, C. (2019). Sentiment analysis of restaurant reviews using machine learning techniques. In V. Sridhar et al. (Eds.), Emerging Research in Electronics, Computer Science and Technology (Vol. 545). Springer. https://doi.org/10.1007/978-981-13-5802-9_60

[18] Yi, S., & Liu, X. (2020). Machine learning-based customer sentiment analysis for recommending shoppers and shops. Complex & Intelligent Systems, 6, 621–634. https://doi.org/10.1007/s40747-020-00155-2

[19] Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for Twitter sentiment analysis. IEEE Access, 6, 23253–23260. https://doi.org/10.1109/access.2017.2776930

[20] Geler, Z., Savić, M., Bratić, B., Kurbalija, V., Ivanović, M., & Dai, W. (2021). Sentiment prediction based on analysis of customers' assessments in food serving businesses. Connection Science, 33(3), 1–19. https://doi.org/10.1080/09540091.2020.1870436

[21] Adak, A., Pradhan, B., & Shukla, N. (2022). Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review. Foods, 11(10), 1500. https://doi.org/10.3390/foods11101500

[22] Gupta, K., Jiwani, N., & Afreen, N. (2023). A combined approach of sentimental analysis using machine learning techniques. Revue d'Intelligence Artificielle, 37(1), 1–6. https://doi.org/10.18280/ria.370101

[23] AlQahtani, A. S. M. (2021). Product sentiment analysis for Amazon reviews. International Journal of Computer Science and Information Technology, 13(3), 15–30. https://doi.org/10.5121/ijcsit.2021.13302

[24]

[25] da Silva, N. F. F., Hruschka, E. R., & Hruschka, E. R. (2014). Tweet sentiment analysis with classifier ensembles. Decision Support Systems, 66, 170–179. https://doi.org/10.1016/j.dss.2014.07.003

[26] Hagen, M., Potthast, M., Büchner, M., & Stein, B. (2015). Twitter sentiment detection via ensemble classification using averaged confidence scores. In Lecture Notes in Computer Science (pp. 741–754). https://doi.org/10.1007/978-3-319-16354-3_81

[27] Saif, H., He, Y., & Alani, H. (2012). Semantic sentiment analysis of Twitter. In ISWC 2012 (pp. 508–524). https://doi.org/10.1007/978-3-642-35176-1_32

[28] Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2013). A knowledge-based approach for polarity classification in Twitter. Journal of the Association for Information Science and Technology, 65(2), 414–425. https://doi.org/10.1002/asi.22984

[29] Paltoglou, G., & Thelwall, M. (2012). Twitter, MySpace, Digg. ACM Transactions on Intelligent Systems and Technology, 3(4), 1–19. https://doi.org/10.1145/2337542.2337551

[30] Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163–173. https://doi.org/10.1002/asi.21662

[31] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. Proceedings of LREC 2010. https://aclanthology.org/L10-1531/

[32] Rana, M. R. R., Nawaz, A., Ali, T., Alattas, A. S., & AbdElminaam, D. S. (2024). Sentiment analysis of product reviews using transformer-enhanced 1D-CNN and BiLSTM. Cybernetics and Information Technologies, 24(3), 112–131. https://doi.org/10.2478/cait- 2024-0028

[33] Ryan, A. A., Arpita, H. D., Tabassum, A., Ahammad, M. S., & Akram, M. A. (2024). A comparative analysis between deep learning and machine learning algorithms based on user review sentiment analysis from various OTT applications. https://doi.org/10.1109/iccece58645.2024.10497229

[34] Pratiwi, R. W., Handayani, S. F., Dairoh, D., & Af'idah, D. I. (2024). Sentiment analysis on Twitter reviews data using the bidirectional long short-term memory (BiLSTM). AIP Conference Proceedings, 3070, 030021. https://doi.org/10.1063/5.0199238

[35] Cheng, Z.-W., & Li, S.-X. (2024). Attention information fusion emotion analysis model based on BiLSTM-BERT. In ICCECE 2024 (pp. 115–119). https://doi.org/10.1109/iccece61317.2024.10504175

[36] Nagelli, A., Saleena, B., & Prakash, B. (2025). Comparative evaluation of deep learning and machine learning techniques for sentiment analysis of electronic product review data. ITM Web of Conferences, 74, 03008. https://doi.org/10.1051/itmconf/20257403008

[37] Kaggle. (2021). Twitter sentiment analysis [Data set]. Kaggle. https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis

221-35-1066

ORIGINALITY REPORT

23%	17%	16%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	thescipub.com Internet Source	3%
3	Submitted to University of Hertfordshire Student Paper	3%
4	Submitted to Daffodil International University Student Paper	2%
5	Submitted to Griffith College Dublin Student Paper	2%
6	ebin.pub Internet Source	1%
7	www.mdpi.com Internet Source	<1%
8	Submitted to The University of the West of Scotland Student Paper	<1%
9	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	<1%
10	Baer, Jing. "A Critical Discourse Analysis: Unveiling the Discursive Strategies of Condom Use Resistance on Reddit's r/TwoXChromosomes", California Institute of Integral Studies	<1%

11	Udara Yedukondalu, V Vijayasri Bolisetty. "Advancing Innovation through AI and Machine Learning Algorithms - Computational Intelligence for Virtual System Optimization. A proceeding of ICMVRCET – 2025", CRC Press, 2025 Publication	<1%
12	"Data Science and Applications", Springer Science and Business Media LLC, 2025 Publication	<1%
13	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
14	Dhirendra Kumar Shukla, Shabir Ali, Sandhya Sharma. "Artificial Intelligence and Sustainable Innovation - Volume 2", CRC Press, 2026 Publication	<1%
15	Manish Kumar Chandan, Shrabanti Mandal. "A comprehensive survey on sentiment analysis: Framework, techniques, and applications", Computer Science Review, 2025 Publication	<1%
16	Submitted to Regent Independent School and Sixth Form College Student Paper	<1%
17	www.cyclopsproject.eu Internet Source	<1%
18	link.springer.com Internet Source	<1%

19 Pijush Dutta, Shobhandeb Paul, Korhan Cengiz, Rishabh Anand, Madhurima Majumder. "A predictive method for emotional sentiment analysis by machine learning from electroencephalography of brainwave data", Elsevier BV, 2023 <1 %
Publication

20 www.warse.org <1 %
Internet Source

21 Mustapha Sahraoui, Laouni Mahmoudi, Mohammed Salem. "An Innovative Arabic Word Embedding Representation for Enhanced Sentiment Analysis", Ingénierie des systèmes d information, 2023 <1 %
Publication

22 "Advanced Computing", Springer Science and Business Media LLC, 2024 <1 %
Publication

23 Rivadeneira Barreiro, Lucia Bernarda. "Decision Modelling Driven by Twitter Data: A Case Study of the 2017 Presidential Election in Ecuador", The University of Manchester (United Kingdom), 2025 <1 %
Publication

24 Ashok Kumar, Geeta Sharma, Anil Sharma, Pooja Chopra, Punam Rattan. "Advances in Networks, Intelligence and Computing - International Conference on Networks, Intelligence and Computing (ICONIC-2023)", CRC Press, 2024 <1 %
Publication

25 John, Chimundu. "Machine Learning Approach for Predictive Maintenance in an <1 %

	Electric Motor; A Classifier Approach.", University of Johannesburg (South Africa) Publication	
26	Varun Sapra, Rohit Tanwar, Luxmi Sapra. "CyberMedics - Navigating AI and Security in the Medical Field", CRC Press, 2025 Publication	<1%
27	Submitted to Liverpool John Moores University Student Paper	<1%
28	doria.fi Internet Source	<1%
29	library.samdu.uz Internet Source	<1%
30	"Intelligent Data Engineering and Automated Learning – IDEAL 2023", Springer Science and Business Media LLC, 2023 Publication	<1%
31	Yahyaeian, Amir Abbas. "Enhancing Mechanical Engineering Education Through a Virtual Instructor in an AI-Driven Virtual Reality Fatigue Test Lab", Purdue University, 2025 Publication	<1%
32	Submitted to University of Hull Student Paper	<1%
33	www.diva-portal.org Internet Source	<1%
34	Submitted to University College Dublin (UCD) Student Paper	<1%
35	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial	<1%

Intelligence, Blockchain, Computing and Security", CRC Press, 2023
Publication

36	arxiv.org Internet Source	<1 %
37	Narendra M. Shekokar, Hari Vasudevan, Surya S. Durbha, Antonis Michalas, Tatwadarshi P. Nagarhalli. "Intelligent Approaches to Cyber Security", Routledge, 2023 Publication	<1 %
38	"Computer Information Systems and Industrial Management", Springer Science and Business Media LLC, 2024 Publication	<1 %
39	ndl.ethernet.edu.et Internet Source	<1 %
40	kth.diva-portal.org Internet Source	<1 %
41	omeka.ibu.edu.ba Internet Source	<1 %
42	saicsit2025.org.za Internet Source	<1 %
43	"Computational Science and Computational Intelligence", Springer Science and Business Media LLC, 2025 Publication	<1 %
44	D. Lakshmi, Ravi Shekhar Tiwari, Rajesh Kumar Dhanaraj, Seifedine Kadry. "Explainable AI (XAI) for Sustainable Development - Trends and Applications", CRC Press, 2024 Publication	<1 %

45	Lestari, Nur Indah. "Optimising Credit Card Fraud Detection Through Machine Learning and Deep Learning with Spatial-Temporal Imbalance Handling", University of Technology Sydney (Australia) Publication	<1 %
46	Suneeta Satpathy, Álvaro Rocha, Sachi Nandan Mohanty, Tanupriya Choudhury. "Intelligent Data-Driven Systems with Innovations in Artificial Intelligence", CRC Press, 2025 Publication	<1 %
47	Submitted to University of Sydney Student Paper	<1 %
48	www.cbit.ac.in Internet Source	<1 %
49	www.preprints.org Internet Source	<1 %
50	Dode, Pragati. "Exploring Classification Methods for Motor Imagery and Execution EEG Signal Fluctuations.", University of Washington Publication	<1 %
51	S. Prasad Jones Christydass, Nurhayati Nurhayati, S. Kannadhasan. "Hybrid and Advanced Technologies", CRC Press, 2025 Publication	<1 %
52	Submitted to University of North Carolina, Greensboro Student Paper	<1 %
53	Keith Cortis, Brian Davis. "Over a decade of social opinion mining: a systematic review", Artificial Intelligence Review, 2021 Publication	<1 %

54	Lakshmi Prasad Mudarakola, Ranjith Kumar Gatla, Akella S. Narasimha Raju, Amar Y. Jaffar et al. "Multi stage sentiment analysis for product reviews on Twitter using optimized machine learning algorithm", Scientific Reports, 2025 Publication	<1%
55	Nidhi Arora, Shilpa Srivastava, Ritu Agarwal, Vandana Mehndiratta, Aprna Tripathi. "Diabetes mellitus prediction using machine learning within the scope of a generic framework", Indonesian Journal of Electrical Engineering and Computer Science, 2023 Publication	<1%
56	Submitted to Ardahan Üniversitesi Student Paper	<1%
57	B. K. Tripathy, Hari Seetha. "Explainable, Interpretable, and Transparent AI Systems", CRC Press, 2024 Publication	<1%
58	Sukhpreet Kaur, Amanpreet Kaur, Manish Kumar. "Recent Advances in Computational Methods in Science and Technology", CRC Press, 2026 Publication	<1%
59	Sushil Kamboj, Pardeep Singh Tiwana. "Innovations in Computing", CRC Press, 2025 Publication	<1%
60	Syarif, Muhammad Fazlurrahman. "Essays on Crowdfunding in Islamic Finance", Hamad Bin Khalifa University (Qatar), 2025 Publication	<1%
61	Usha Desai, Biswaranjan Acharya, Madhu Shukla, Varadraj Gurupur. "Decision Support	<1%

System for Diabetes Healthcare: Advancements and Applications", River Publishers, 2025
Publication

62	Gyan Prakash, Amandeep Kaur. "AI and Sustainable Transformations", CRC Press, 2025 Publication	<1 %
63	researchinnovationjournal.com Internet Source	<1 %
64	Hesham Alhumyani, Ibrahim Alrube, Sameer Alsharif, Ashraf Afifi, Chokri Ben Amar, Hala S. El-Sayed, Osama S. Faragallah. "An Efficient Internet Traffic Classification System Using Deep Learning for IoT", Computers, Materials & Continua, 2022 Publication	<1 %
65	Md. Saiful Islam, Arafatun Noor Orno, Mohammad Arifuzzaman. "Approach to Social Media Cyberbullying and Harassment Detection Using Advanced Machine Learning", Springer Science and Business Media LLC, 2024 Publication	<1 %
66	Submitted to University of Exeter Student Paper	<1 %
67	bmcpublichealth.biomedcentral.com Internet Source	<1 %
68	pdfs.semanticscholar.org Internet Source	<1 %
69	www.sciencegate.app Internet Source	<1 %
70	yorkspace.library.yorku.ca Internet Source	

<1 %

71	Alboom, Rashed Khalid. "Energy Forecasting Inaccuracies and Their Direct Impact on Grid Performance", Rochester Institute of Technology Publication	<1 %
72	Michael Mensah[1], Sampson Opoku[2], Selasie Brown[3], Prince Afriyie[4]. "Predictive Models and Predictors of Under-5 Mortality Using Machine Learning Techniques: Evidence from the 2022 DHS", Springer Science and Business Media LLC, 2025 Publication	<1 %
73	Nwokolo, Ijeoma Jacinta. "Harnessing Machine Learning to Predict At-Risk Students: A Data-Driven Approach for Early Intervention", Edge Hill University (United Kingdom) Publication	<1 %
74	Razieh Abedi Rad, Mohammad Reza Yamaghani, Azamossadat Nourbakhsh. "A survey of sentiment analysis methods based on graph neural network", International Journal of Data Science and Analytics, 2025 Publication	<1 %
75	www.ijraset.com Internet Source	<1 %
76	www.jist.ir Internet Source	<1 %
77	Amit Kumar Tyagi. "Data Science and Data Analytics - Opportunities and Challenges", CRC Press, 2021 Publication	<1 %

78	Arvind Dagur, Sohit Agarwal, Dhirendra Kumar Shukla, Shabir Ali, Sandhya Sharma. "Artificial Intelligence and Sustainable Innovation - Volume 1", CRC Press, 2026 Publication	<1 %
79	Küp, Büşra Ülkü. "Real Time Prediction of Delivery Delay With Machine Learning", Kadir Has Üniversitesi Publication	<1 %
80	Praveen Kumar Shukla, Harshal A. Sanghvi, Ankur Agarwal, Abhijit S. Pandya. "Artificial Intelligence and Internet of Things for Smart Agriculture", CRC Press, 2025 Publication	<1 %
81	Xin Ye, Hongxia Dai, Lu-an Dong, Xinyue Wang. "Multi-view ensemble learning method for microblog sentiment classification", Expert Systems with Applications, 2021 Publication	<1 %
82	dergipark.org.tr Internet Source	<1 %
83	newinera.com Internet Source	<1 %
84	researchspace.ukzn.ac.za Internet Source	<1 %
85	www.jnao-nu.com Internet Source	<1 %
86	www.techscience.com Internet Source	<1 %
87	"Mathematical Modeling, Computational Intelligence Techniques and Renewable Energy", Springer Science and Business Media LLC, 2025	<1 %

88	Aashita Chhabra, Mansaf Alam. "RegGRU-Opt: A Robust GRU-Based RNN Model for High-Performance Sentiment Analysis and Binary Classification", Arabian Journal for Science and Engineering, 2025 Publication	<1 %
89	Alabi, Ibrahim Olalekan. "Advancing Snow Water Equivalent Monitoring With Machine Learning and L-Band Interferometric Synthetic Aperture Radar (InSAR) Data", Boise State University, 2025 Publication	<1 %
90	Andi Muhammad Iqra Rezky Hatta, Lilis Nur Hayati, Herdianti Herdianti. "Performance Analysis of Gradient Boosting and Decision Tree on Distributed Denial of Service Attacks in Software Defined Networks", Jurnal Sisfokom (Sistem Informasi dan Komputer), 2025 Publication	<1 %
91	C Kishor Kumar Reddy, Anindya Nag, Lavanya Pamulaparty, Mariya Ouaissa, Marlia Mohd Hanafiah. "Generative AI in Neurology - Advancing Neurodegenerative Disease Treatment", CRC Press, 2025 Publication	<1 %
92	Gaurav Jaswal, Vivek Kanhangad, Raghavendra Ramachandra. "AI and Deep Learning in Biometric Security - Trends, Potential, and Challenges", CRC Press, 2021 Publication	<1 %
93	Katsarou, Aikaterini. "Sequential Machine Learning For Textual and Time-Series Data",	<1 %

Technische Universitaet Berlin (Germany), 2025
Publication

94	María Reina Santiago-Rosario, Asha Yadav, Kent McIntosh, Haley Cerdan. "Measuring teacher emotional language valence via sentiment analysis: Correlations with classroom practices", Journal of School Psychology, 2026 Publication	<1 %
95	Mechelen Bruxellas, Brynna. "Intelligent Turfgrass Evaluation: Leveraging AI and UAVs for Automated Phenotypic Assessment", North Carolina State University, 2025 Publication	<1 %
96	Sachin Deori, Tulika Chutia, Hirok Jyoti Saikia, Anjan Boro, Nomi Baruah. "Bridging techniques and applications in sentiment analysis: Approaches, challenges, and emerging trends", Engineering Applications of Artificial Intelligence, 2026 Publication	<1 %
97	Shivam Sharma, Rahul Aggarwal, Manoj Kumar. "Mining Twitter for Insights into ChatGPT Sentiment: A Machine Learning Approach", 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2023 Publication	<1 %
98	Thapa, Samir. "Improving Detection of Faults and Cybersecurity in Power System Using Digital Twin and SVM.", Purdue University Publication	<1 %
99	content.wolfram.com Internet Source	<1 %

100	de Carvalho, Isabel Maria Rocha Correia. "Sentiment Analysis in the Context of Dialogue", Universidade de Coimbra (Portugal), 2024 Publication	<1 %
101	eprints.leedsbeckett.ac.uk Internet Source	<1 %
102	era.library.ualberta.ca Internet Source	<1 %
103	files.core.ac.uk Internet Source	<1 %
104	fourweekmba.com Internet Source	<1 %
105	ir.canterbury.ac.nz Internet Source	<1 %
106	jisem-journal.com Internet Source	<1 %
107	journals.openedition.org Internet Source	<1 %
108	ntnuopen.ntnu.no Internet Source	<1 %
109	nuleren.be Internet Source	<1 %
110	thesai.org Internet Source	<1 %
111	umu.diva-portal.org Internet Source	<1 %
112	www.frontiersin.org Internet Source	<1 %
113	www.iieta.org Internet Source	<1 %

114	www.ozguryayinlari.com Internet Source	<1 %
115	www.tnsroindia.org.in Internet Source	<1 %
116	"Data Intelligence and Cognitive Informatics", Springer Science and Business Media LLC, 2021 Publication	<1 %
117	Jamal Mabrouki. "Integration of Advanced Systems in Environmental Science and Water Desalination", CRC Press, 2026 Publication	<1 %
118	"Advanced Informatics for Computing Research", Springer Science and Business Media LLC, 2025 Publication	<1 %
119	"Computer Networks and Inventive Communication Technologies", Springer Science and Business Media LLC, 2023 Publication	<1 %
120	Amit Kumar Tyagi, Shrikant Tiwari, S. V. Nagaraj. "Quantum Computing - The Future of Information Processing", CRC Press, 2025 Publication	<1 %
121	Amit Lathigara, Nirav Bhatt, Paresh Tanna. "Emerging Perspectives and Applications of Computational Intelligence and Smart Systems - Proceedings of the International Conference on Computational Intelligence, Emerging Technologies, and Smart Systems (ICCIETSS 2025), March 28-29, 2025, Rajkot, Gujarat, India", CRC Press, 2025 Publication	<1 %

122	Kim, Eunyoung. "A Comparative Study of Clustering and Bert-Based Classification Methods for Short-Text Reflections Analysis.", The University of North Carolina at Charlotte Publication	<1%
123	Rahul Kumar Chaurasiya, Varun Bajaj, Vishakha Chourasia. "Artificial Intelligence-based Signal Processing for Brain Activity Analysis", CRC Press, 2025 Publication	<1%
124	Urmila Shrawankar, Prerna Mishra. "Cloud Computing for Smart Education and Collaborative Learning", Routledge, 2025 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

221-35-1066

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64

PAGE 65

PAGE 66

PAGE 67

PAGE 68

PAGE 69

PAGE 72

PAGE 73

PAGE 74

PAGE 75

PAGE 76

PAGE 77

Account Clearance