A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF ÇANKIRI KARATEKİN UNIVERSITY

USER BEHAVIOR ANALYSIS ON E-COMMERCE USING NLP TECHNIQUES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRONICS AND COMPUTER ENGINEERING

BY

ASMAA SAMI MIRDAN MIRDAN

ÇANKIRI

2023

USER BEHAVIOR ANALYSIS ON E-COMMERCE USING NLP TECHNIQUES

By Asmaa Sami Mirdan MIRDAN

July 2023

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science

Advisor **:** Asst. Prof. Dr. Selim BUYRUKOĞLU
Co-Advisor **:** Asst. Prof. Dr. Mohammed Rashad Baker BAKER

Examining Committee Members:

Chairman **:** Asst. Prof. Dr. Selim BUYRUKOĞLU
Electronics and Computer Engineering
Çankırı Karatekin University

Member **:** Asst. Prof. Dr. Fuat TÜRK
Electronics and Computer Engineering
Kırıkkale University

Member **:** Asst. Prof. Dr. Mustafa KARAHAN
Electronics and Computer Engineering
Çankırı Karatekin University

Approved for the Graduate School of Natural and Applied Sciences

Prof. Dr. Hamit ALYAR
Director of Graduate School

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Asmaa Sami Mirdan MIRDAN

ABSTRACT

USER BEHAVIOR ANALYSIS ON E-COMMERCE USING NLP TECHNIQUES

Asmaa Sami Mirdan MIRDAN
Master of Science in Electronics and Computer Engineering
Advisor: Asst. Prof. Dr. Selim BUYRUKOĞLU
Co-Advisor: Asst. Prof. Dr. Mohammed Rashad Baker BAKER
July 2023

This study presents an in-depth investigation into the potential for sentiment analysis (SA) and machine learning (ML) in facilitating the sales prediction and customer retention processes for both small and large-scale businesses. Online platforms including blogs, social networks, and review portals have transformed the marketing landscape, enabling consumers to voice their opinions on a vast array of topics, from product reviews to popular culture. These digital platforms not only foster customer engagement, but also provide businesses with an invaluable data source for predictive analysis, essential in strategic sales forecasting and customer relationship management. In this study, we compiled a comprehensive dataset of product review tweets, serving as a rich representation of consumer sentiment. To ensure the integrity and relevance of the data, we engaged in rigorous preprocessing methodologies, mitigating potential noise and inconsistencies. Following the cleaning phase, we utilized the Valence Aware Dictionary and sEntiment Reasoner (VADER), a lexicon and rule-based sentiment analysis tool, in conjunction with several machine learning algorithms. The objective was to ascertain the most effective means of classifying the sentiment polarity of product reviews, subsequently aiding in sales prediction. Our findings reveal that while VADER offers notable benefits in sentiment analysis, ML techniques present superior accuracy in classifying the polarity of product reviews. More specifically, logistic regression (LR) was found to be the top-performing algorithm in this context. Across a multitude of evaluation metrics, including accuracy, precision, recall, F1-score, Matthews correlation coefficient (MCC), and area under the curve (AUC), LR consistently outperformed its counterparts, thus solidifying its position as the most apt choice for sentiment classification in product reviews. Notably, other algorithms such as XGBoost and Stochastic Gradient Descent (SGD) also demonstrated competitive performance. They can serve as plausible alternatives in situations where model interpretability is not the prime concern and a higher degree of model complexity is permissible. These findings contribute to an emerging body of knowledge, illuminating the potential of SA and ML in providing businesses with robust tools for understanding customer sentiment, predicting sales, and consequently enhancing customer retention strategies. The implications of this study extend beyond academia, promising substantial real-world benefits for various stakeholders in the business sphere.

2023, 62 pages

Keywords: Unbalanced dataset, Sentiment analysis, Machine learning, Classification, Product review

ÖZET

NLP TEKNİKLERİ KULLANARAK E-TİCARETTE KULLANICI DAVRANIŞI ANALİZİ

Asmaa Sami Mirdan MIRDAN
Elektronik ve Bilgisayar Mühendisliği, Yüksek Lisans
Tez Danışmanı: Dr. Öğr. Üyesi Selim BUYRUKOĞLU
Eş Danışman: Dr. Öğr. Üyesi Mohammed Rashad Baker BAKER
Temmuz 2023

Bu çalışma, küçük ve büyük ölçekli işletmeler için satış tahminlerini ve müşteri sadakatini kolaylaştırmada duygu analizi (SA) ve makine öğrenmesi (ML) potansiyelini derinlemesine bir inceleme sunmaktadır. Bloglar, sosyal ağlar ve inceleme portalları dahil olmak üzere çevrimiçi platformlar, pazarlama manzarasını dönüştürmüş, tüketicilerin ürün incelemelerinden popüler kültüre kadar çok çeşitli konularda fikirlerini ifade etmelerini mümkün kılmıştır. Bu dijital platformlar sadece müşteri katılımını teşvik etmekle kalmaz, aynı zamanda işletmelere stratejik satış tahminleri ve müşteri ilişkileri yönetiminde hayati önem taşıyan tahmini analizler için paha biçilmez bir veri kaynağı sağlar. Bu çalışmada, tüketici duygusunun zengin bir temsili olan kapsamlı bir ürün inceleme tweetleri veri seti derledik. Verinin bütünlüğünü ve alakasını sağlamak için, potansiyel gürültü ve tutarsızlıkları hafifleten sıkı ön işleme metodolojilerine başvurduk. Temizleme aşamasını takiben, bir lexicon ve kural tabanlı duygu analiz aracı olan Valence Aware Dictionary and sEntiment Reasoner (VADER)'i bir dizi makine öğrenmesi algoritmasıyla birlikte kullandık. Hedefimiz, satış tahminine yardımcı olmak için ürün incelemelerinin duygu kutuplaşmasını sınıflandırmanın en etkili yolunu belirlemekti. Bulgularımız, VADER'ın duygu analizinde dikkate değer avantajlar sunduğunu, ancak ML tekniklerinin ürün incelemelerinin kutuplaşmasını sınıflandırmada daha üstün doğruluk sunduğunu ortaya koymaktadır. Daha spesifik olarak, lojistik regresyon (LR), bu bağlamda en iyi performans gösteren algoritma olarak bulunmuştur. Doğruluk, hassasiyet, geri çağırma, F1-skoru, Matthews korelasyon katsayısı (MCC) ve eğri altındaki alan (AUC) dahil olmak üzere çok sayıda değerlendirme metriğinde, LR sürekli olarak diğer algoritmalardan daha üstün performans sergileyerek, ürün incelemelerinde duygu sınıflandırması için en uygun seçenek olarak konumunu sağlamlaştırmıştır. Dikkat çekici şekilde, XGBoost ve Stochastic Gradient Descent (SGD) gibi diğer algoritmalar da rekabetçi performans sergilemiştir. Model yorumlanabilirliği ana endişe olmadığı ve daha yüksek model karmaşıklığına izin verilebilecek durumlarda makul alternatifler olarak hizmet edebilirler. Bu bulgular, SA ve ML'in işletmelere, müşteri duygusunu anlama, satışları tahmin etme ve sonuç olarak müşteri sadakat stratejilerini güçlendirme konusunda sağlam araçlar sağlama potansiyelini aydınlatan yeni bir bilgi birikimine katkıda bulunur. Bu çalışmanın sonuçları, iş dünyasındaki çeşitli paydaşlar için önemli gerçek dünya faydaları vaat eden akademi ötesine uzanmaktadır.

2023, 62 sayfa

Anahtar Kelimeler: Dengesiz veri seti, Duygu analizi, Makine öğrenmesi, Sınıflandırma, Ürün incelemesi

PREFACE AND ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Asst. Prof. Dr. Selim BUYRUKOĞLU and Asst. Prof. Dr. Mohammed Rashad Baker BAKER, for their patience, guidance and understanding.

Asmaa Sami Mirdan MIRDAN
Çankırı-2023

CONTENTS

LIST OF ABBREVIATIONS

AdaBoost	Adaptive boosting
XGBoost	Extreme gradient boosting
LR	Logistic regression
ML	Machine learning
NLP	Natural language processing
SA	Sentiment analysis
SGD	Stochastic gradient descent
SMOTE	Synthetic minority oversampling technique
TF-IDF	Term frequency-inverse document frequency
VADER	Valence aware dictionary and entiment reasoner
MCC	Matthews correlation coefficient
AUC	Area under curve

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

E-commerce has revolutionized the way people shop and for many has become an essential part of everyday life (Gunasekaran *et al.* 2002). With the rapid growth of e-commerce, understanding user behavior has become crucial for businesses to provide a personalized and seamless experience to their customers. Most of the time, a consumer buys or cancels a product based on reviews alone. Consequently, it is clear that surveys are beneficial (Fisher and Kordupleski 2019). However, it can be challenging to sift through hundreds of reviews every time someone is considering buying a product (Hu and Liu 2004). As a result, it will be beneficial to extract some relevant information from these evaluations. On the other hand, every business organization depends on the intelligent decision-making analytical system to analyze consumer behavior (Alasiri and Salameh 2020). The demand-based supply chain management in the organization may be most significantly impacted by this analysis and prediction. Data analysts employ a variety of tools, including machine learning (ML) techniques and data mining, to find hidden patterns in consumer behavior and forecast sales. The newest strategy employing data mining and ML has paved the best road to uncover the hidden layers because traditional methods of analysis cannot match the speed of data generated by existing e-commerce sites (Micol Policarpo *et al.* 2021). Now everything has been changed by ML and artificial intelligence (AI). Applications of ML have been widely researched in areas such as business (Baker *et al.* 2022), sentiment analysis (SA) (A. Alamoodi *et al.* 2020), etc. Natural language processing (NLP) techniques and ML algorithms are effective in analyzing user behavior in e-commerce platforms. Using these technologies, businesses can gain insight into customer preferences, buying patterns and other valuable information that can help them optimize marketing strategies, improve customer satisfaction and increase revenue. In this thesis, we examine the application of NLP techniques and ML algorithms in the analysis of user behavior in e-commerce platforms and discuss their advantages and limitations which all seek to satisfy the customer and keep them coming back to a specific online store. The importance of the present research has been examined from two theoretical and scientific perspectives and its necessity in two spatial and temporal dimensions, which will be explained below.

1.1 Importance of Thesis from a Theoretical Point of View

The widespread development of technology in Internet-connected devices and services has led to a significant increase in electronic commerce. Modern devices and advanced applications have become widely available to people, driving the trend towards online shopping. As e-commerce expands, numerous companies and service centers utilize the web to offer their products and services.

The rise of online shopping has attracted a growing number of Internet users who prefer to purchase products and services online (O'Cass and Fenech 2003). However, the vast array of products available on e-commerce websites can sometimes overwhelm customers and make it challenging for them to find the right product. This high level of competition among global trading sites emphasizes the need for effective strategies to increase financial profit (London and Hart 2004).

In today's age, people register their opinions on various reviews of related products, brands and services on the internet and exchange opinions with other people (Liao *et al.* 2021). The opinions generated are valuable assets that can be used to inform important decisions. In this way, checking opinions related to products and services can not only improve their quality, but also influence the purchasing decisions of users. Vast of people's purchases; through getting to know products and services in the virtual space; and based on comments provided by other users. Users and customers pay a lot of attention to information that includes other people's feelings (Rabjohn *et al.* 2008). Because some people do not understand technical and professional concepts well, while they understand words like good and bad (Dong and Jiang 2019).

On the other hand, many large companies use customer reviews from online shopping sites and other web pages to develop and improve their business, including customer relationship management (CRM), increasing customer satisfaction, customer retention, and sales (Zeng *et al.* 2003). They also use them to build their reputation and build brand awareness. Businesses and organizations have spent a lot of money on studies and consultants to find out what consumers think and feel about their goods and services. This is despite the fact that the text comments on the Internet are easily and freely available, and companies and organizations can support their products and services by using survey techniques and comment analysis and take steps towards the success of the organization.

Based on this, SA of users' opinions is considered one of the research fields of interest for researchers in the market field (Medhat *et al.* 2014). Currently, with the increase in the number of people who use websites and other social media to express their opinions, the ability to perform SA in order to predict people's opinions and attitudes has also increased. SA is the robotic exploration of opinions and feelings through NLP, which involves classifying opinions into positive, negative, or neutral groups. However, consumer behavior itself is a complex pattern among society. With the aim of predicting the likelihood of such patterns, the researchers applied several probabilistic statistical data mining approaches and ML models to historical online customer data. This resulted in somewhat reliable probabilities for predicting the customer's next steps (Cirqueira *et al.* 2020).

The focus of this system is to improve the performance of e-commerce systems by facilitating customers to find the right products according to their preferences. In this thesis, a new conceptual framework for analyzing and predicting customer steps in online shopping is proposed. This thesis first draws the existing literature about the adopted data sets, forecasting methods and tasks with their applications. Then a system based on customers' behavior and collaboration with NLP and ML to support customers' decision making is proposed.

1.2 Importance of Thesis from a Scientific Point of View

The use of online tools and technologies for various purposes by 21st century companies has increased thanks to the excellent access to the Internet (Berisha-Shaqiri and Berisha-Namani 2015). The development of information and communication technology has provided a basis for people around the world to benefit from the products and services of companies and organizations. Therefore, according to the current trend, many customers prefer online shopping. In this regard, users try to find out the opinions of other buyers and consumers regarding products and services and finally make the right decision about buying a product or service. On the other hand, market analysts and sales managers need to understand the behavior of consumers and the factors that encourage them to buy and analyze the sentiments of users in order to maintain their position in the online competition system.

Therefore, the problem of analyzing customers' emotions and behavior in online shopping is a growing research topic that has its own difficulty and complexity (Smith and Sivakumar 2004). On the other hand, social media, which are important sources of data for SA, are constantly expanding and produce much more complex and relevant information (Medhat *et al.* 2014). Although SA is an important area and currently has a wide range of applications, it is clearly not a simple task and there are many technical challenges associated with NLP. Therefore, the efficiency and accuracy of SA face the challenges faced in NLP, and data mining techniques have become the focus of such analysis (Khan *et al.* 2016).

1.3 Research Innovation

In the increasingly interconnected and digital world, the vast amounts of data generated on social media platforms, such as Twitter, have become a treasure trove of insights. Among these, understanding and quantifying public sentiment stands as a pivotal component for various fields, ranging from market research to public policy. The present study is situated within this context, aiming to delve into the landscape of public sentiment using a robust methodology. The study embarked on the task of collating a significant dataset from Twitter, one of the most vibrant and information-rich social media platforms. These tweets serve as candid reflections of public sentiment, which, when analyzed systematically, can offer compelling insights. To this end, the gathered data was meticulously categorized using a lexicon - a dictionary of words categorized by sentiment polarity - effectively transforming raw tweets into a structured dataset. The heart of this approach lies in the classification of sentiments within the lexicon as either positive or negative. After the data categorization, the study derived the proportion of positive and negative sentiments from the lexicon, which was utilized to represent the overall sentiment landscape.. People who use a social network or social media sites such as Facebook or Twitter express and share their opinions on specific topics such as news, movies, events, or purchasing a specific product (Mehta *et al.* 2021). SA has become a hot research topic in the field of NLP due to the important role it plays in analyzing public opinion and inferring opinion-based decisions. SA is a field of study in which people's views and expressions are classified as positive, negative, or neutral. Several definitions are found in the literature, but SA is best defined as analytics used to extract data based on user sentiment. Sentiment analysis, also known as opinion mining, is the study of opinions, thoughts, experiences, feelings, and behaviors in the form of text (Alamoodi *et al.* 2022).

In this regard, this research wants to help understand the behavior of buyers and sellers in the e-commerce platform by analyzing the feelings of customers and users, and in this regard, the following items are considered among the innovations of the research:

- Using the keyword technique to display the words that have the most weight and the most important and most frequent words used by users about products and services in the comments text. The purpose of this stage is that the audience, users and customers can understand the opinions of other buyers in a more concrete way.
- Finally, the use of Extreme gradient boosting (XGBoost) algorithm to classify sentiments in the text of opinions is considered as the most important and novel innovation in this research, because other researches in the field of opinion mining use other ML and collective learning algorithms for sentiment analysis.

1.4 Research Challenges

Today, social media has become a great source for getting users' opinions about products, services, and various topics. Blogs and websites are a real-time tool for gathering product feedback. However, the overabundance of blogs in the cloud has generated a huge amount of information in various forms such as opinions, comments and reviews. Therefore, there is an urgent need to find a method to extract meaningful information from big data, classify it into different categories and predict the behavior or emotions of the end user. But some of the major challenges in SA are:

1) A number of comments have positive sentiments in one situation and negative sentiments in another situation.
2) People do not share their same opinions.
3) Users mostly use informal and colloquial sentences, abbreviations and emoticons to express their opinions.
4) Users do not express their feelings directly and may use negative verbs or adjectives, sarcasm, sarcasm or jokes in expressing their opinions, each of which creates a concept completely different from the real meaning of the opinion. Therefore, these sentences are considered as special sentences and the words should be checked carefully.
5) Interrogative and conditional sentences may not have a positive or negative feeling, but the keyword used in it may be positive or negative.
6) Sometimes there may be no emotional words such as good, better, excellent, bad, worst, etc. in the sentences. However, sentences may have positive and negative feedback.
7) The large volume of words in natural language is considered another complication of analysis.

The existence of spam comments and their identification from the text of comments provided by real users and customers is difficult and is considered one of the important challenges in this field. Spam comments are those comments that are provided by the organization, company or even some people to increase the value of the organization or the product and services of the relevant company in order to attract more users and customers.

1.5 Research Objectives

There are many goals in this research, among which the main goals are:

- Providing an innovative model for predicting the behavior of customers and consumers of online markets based on NLP methods and ML algorithms.
- Providing a better performance method for analyzing customer sentiments in social media networks on multiple sizes of Twitter database.
- Data preprocessing and use of VADER SA and ML algorithms to classify positive or negative user opinions about online purchases.
- Improving the accuracy parameter of the classification of product comments in the proposed method compared to the previous methods

1.6 Research Structure

The thesis structure will be organized and presented as follows:

- Second section: This part includes two parts of literature and research background, and it includes discussion and review related to the basic and infrastructural concepts of the research, as well as the analysis and description of the details of the research records.
- Third section: In this part, the proposed method will be introduced and key details related to how it will be designed and modeled in order to achieve the research objectives will be presented.
- Fourth section: This part includes the topics related to the simulation of the proposed method and its comparison with past researches, and the purpose of its presentation is to evaluate the efficiency of the proposed method and examine its capabilities and limitations in competition with past researches.
- Fifth section: This part includes general conclusions from the results of the research and suggestions for future work.

2. LITERATURE REVIEW

Rapid developments in the use of the Internet have created a need for predictive analytics to help understand the behavior of buyers and sellers on e-commerce platforms. With the development of the e-commerce platform, satisfactory products have been provided for customers and customer loyalty has increased (Nisar and Prabhakar 2017). Online shopping has evolved over the years since its inception and has also inherited many advancements in lifestyle, shopping and business. Finding new potential buyers, retargeting existing buyer, handling inquiries and maintaining inventory are very challenging modules and if not taken care of, it will affect the e-commerce industry badly. In this regard, in this section, data mining methods, NLP, and user SA in social networks such as Twitter and predicting customer behavior in online shopping are discussed.

In recent years, due to the growth of the Internet space, the sharing of Sentiment by users on social networks, blogs, product review pages, as well as websites for online sales of products and services, has increased dramatically (Sin *et al.* 2012). This issue has led to the emergence of a new field of work in the science of NLP called opinion mining. Opinion research, also known as SA, analyzes people's Sentiments, feelings, attitudes, and opinions about products, services, organizations, people, issues, events, topics, and their characteristics. The process of opinion mining, as one of the most important modern approaches in data mining, includes understanding and extracting human emotions from simple and unstructured textual data. In fact, opinion analysis is a sub-branch of text analysis that analyzes and examines a written language with the main purpose of extracting theoretical, emotional expressions, expressed tendencies, and emotions (Can and Alatas 2019). Based on this, opinion polling can be applied to various areas such as examining customer opinions, business intelligence, marketing promotion, rating services, and products, analyzing political and sports opinions, transactions, social media, and interactions to extract trends and opinions (Fan and Gordon 2014).

In recent years, due to the rapid growth of e-commerce and the increasing importance of the competitive business environment, researchers are trying to use more and more opinion polling techniques in order to analyze sentiments and summarize the opinions of users and customers. The analysis of users' opinions about products and services plays a crucial role in assisting other customers in making purchasing decisions, thereby reducing doubts. It also aids manufacturers and traders in refining their marketing strategies for products and services. Consequently, users have become the primary audience for opinion and SA, with nearly 15% of customers checking online reviews before making a purchase. "What do others think?" contains vital information in the decision-making process for potential buyers. Moreover, besides organizations, producers, company owners, and service providers are also involved in this field. Some researchers consider opinion polling as a ML approach in which machines analyze sentiments, inclinations, and opinions expressed by people through text, star ratings, etc., and classify them accordingly (Gupta *et al.* 2020).

SA, also known as opinion mining, is a NLP technique that identifies and extracts subjective information from text data. The goal is to determine the attitude, opinions, or emotions expressed in a piece of text, especially to ascertain whether the writer's attitude towards a particular topic is positive, negative, or neutral (Liu 2012).

SA has become an increasingly active research area in recent years due to its wide range of applications. It has been applied extensively to analyze social media data, product reviews, political discourse, and financial news (Medhat *et al.* 2014). With the explosive growth of social media and user-generated content on the web, SA has emerged as a key technique to understand opinions and emotions at scale (Cambria *et al.* 2013). Major companies now routinely apply SA to analyze customer feedback, monitor brand and product perceptions, and understand public attitudes.

The early work in SA focused on document-level classification, categorizing whole documents as expressing positive or negative sentiment (Pang *et al.* 2002). ML methods like support vector machines and naive Bayes were applied on document feature vectors. However, this ignored sentiment towards specific entities and topics. Later research moved towards more granular aspect-based SA (Pontiki *et al.* 2014). This involves extracting opinion targets and sentiment polarity towards each target. For example, a review of a smartphone may express positive sentiment towards screen quality but negative sentiment towards battery life.

Aspect-based SA aims to extract fine-grained information that is more useful for many applications. A variety of supervised and unsupervised methods have been explored for aspect extraction and sentiment classification (Zhang *et al.* 2018). Conditional random fields, recurrent neural networks, and attention mechanisms have shown success by incorporating context and syntax information. Weakly supervised and distant supervision techniques help reduce the dependency on large labeled datasets. But performance remains significantly lower than human capability.

Sarcasm, irony, metaphor, ambiguity, domain dependence, and context dependence remain key challenges (Liu 2012). Sentiment is often expressed in nuanced ways that need deeper linguistic understanding. Recent work has applied linguistic knowledge resources like SenticNet and Sentic LSTM to better handle concept-level semantics (Cambria *et al.* 2018). But accurately modeling the complexity of language understanding remains an open research problem.

Overall, SA has grown rapidly with a wide array of applications. But significant gaps remain compared to human-level language capabilities (Hovy and Yang 2021). Key opportunities for future work include improving model generalization across domains, better incorporating linguistic knowledge, reducing dependency on labeled data through semi-supervised learning, and multimodal SA combining text, audio and visual inputs.

Table 2.1 shows the results of this research in relation to the comparison of three analysis approaches, based on which it is determined that there is no need for a dictionary in the ML approach and among the disadvantages of this approach is the domain dependency. While the dictionary-based approach does not require labeled data, it requires a rich dictionary that may not always be available.

Table 2.1 Comparison of analysis approaches

APPROACH	CLASSIFICATION	ADVANTAGES	DISADVANTAGES
ML based approach	Learning with and without a supervisor	A dictionary is not necessary. Show high classification accuracy	Text classification in one trained domain does not work with other domains in most cases
Law-based approach	Supervised and unsupervised learning	Performance accuracy is 11% at the review level and 16% at the sentence level. In the classification of emotions The sentence level has better performance than the word level.	Efficiency and accuracy depend on defined rules.
A vocabulary-based approach	Unsupervised learning	Labeled data and learning method are not required	The need for strong language resources is not always available

2.1 Natural Language Processing and Text Mining

NLP and text mining are closely related fields that focus on extracting meaning and insights from human language. NLP involves the use of computational techniques and algorithms to understand, analyze, and generate human language. It encompasses various tasks such as speech recognition, language translation, SA, and information retrieval.

Text mining, on the other hand, specifically deals with extracting valuable information and knowledge from large collections of textual data. It involves applying statistical and ML techniques to analyze patterns, relationships, and trends within the text. Text mining can be used for tasks such as text classification, topic modeling, entity extraction, and summarization.

Both NLP and text mining rely on advanced algorithms and models to process and interpret human language. They leverage techniques such as ML, deep learning, and linguistic analysis to extract meaningful insights from text data. These fields have applications in various domains, including customer feedback analysis, social media mining, market research, information retrieval, and automated content generation (Kumar and Sebastian 2012).

2.1.1 Applications of natural language processing

NLP is used in two areas: speech and writing. Written applications include text translation, SA, summarizing texts, and so on. Voice assistants and robots that respond to human speech are among its speech applications (Kumar and Sebastian 2012).

In general, the applications of NLP in our daily life, educational environments, therapy and industry can be divided into several categories as follows:

- Speech processing
- Image processing
- Text processing

1) Applications of speech processing

- Personal assistants: The purpose of these smart programs is to execute human voice commands on a smartphone or computer. Such as: Apple Siri, Amazon Alexa, Google Assistant, Microsoft Cortana, Samsung Bixby and Raymon Persian voice assistant.
- Speech to text conversion: The purpose of this tool is to convert sound to text equivalent to the speaker's words. For Farsi language, it is recommended to use the Google keyboard for smartphones and Navisa software. Other well-known speech-to-text conversion engines are wit.ai, Google Speech, and Yandex SpeechKit.
- Text-to-speech conversion: just unlike the previous tool, here the goal is to convert text to speech. This tool is used to read the text of the message by voice in many software and queue systems, etc. Such as: Ariana Persian software.
- Voice translators (Translator): who translate the speaker's voice from one language to another. Like: Google Translate.
- Other applications of speech processing are less in the subfield of NLP science. For other applications of speech processing, we can refer to its use in the telecommunications and communication industry, recognizing the person speaking

from the voice (for security applications), recognizing the sense of the speaker (or the level of truthfulness of the speaker), etc.

2) Image processing applications

- Optical Character Recognition/Reader or OCR for short: it is the automatic recognition of texts in document images and converting them into searchable and editable texts by computer.
- Image translators: who extract the text from the image and translate it into another language. Like: Google Translate.
- Image Captioning: explaining the elements and events in the image, which is usually done by deep learning techniques. For example: receiving an output image, it writes that "two birds are sitting on a tree branch".

3) Text processing applications (text mining applications) and information retrieval

- Opinion analysis (sensory analysis of the text): This involves analyzing the general level of satisfaction of people, users or customers in elections (and other political issues), stores, service centers, etc.
- Automatic summarization: This refers to reducing the volume of the text while maintaining the concepts (subject) and continuity (readability) of the original text. It has applications such as news summary, site content summary as a result of search engines.
- Machine translation of the text: This involves translating sentences in the text from the source language to the target language.
- Similarity and disambiguation of words in the text: This task deals with identifying similar words and clarifying ambiguous ones.
- Detection of literary (scientific) fraud: This refers to the identification of any fraudulent activities in the text.
- Identifying the author (or the author's gender) of the text based on the writing style: This task uses stylistic cues to determine the author or the author's gender.

- Language generation: This involves converting or expressing the information in the database into human language by the machine.
- Text enrichment: This refers to annotation, and added value in text for search engines and other textual semantic analysis.
- Question and answer systems and chatbots: These involve direct interactions with users to provide information or assistance.
- Search engine: This entails the production and optimization of various components of search engines for high volumes of data.
- Information extraction: This involves discovering entities and relationships between them in the text.
- Keyword extraction: This is aimed at tagging or automatic topic tagging of the text.
- Classification and clustering of texts: This involves grouping (supervised or unsupervised) sets of texts with applications such as recognizing the topic of the text (any group of texts), automatic indexing of the text, grouping similar texts (news) with the aim of identifying important issues/events/... in the mass Texts of social networks or news.
- Basic NLP tools: These include lexical networks, parsers, semantic tagging of words, pronoun reference discovery, recognition and classification of nouns.
- The basic operation of information retrieval: This involves transforming text into numerical vectors, determining distance criteria or textual similarity, and feature engineering

2.2 Sentiment Analysis

SA is focused on the analysis of whether a person's feelings or words are positive, negative, or neutral. While there are several different definitions of SA in the literature, the most precise one suggests that it is analytics applied to the extraction of data based on user sentiment. Also known as opinion mining, SA studies people's written expressions of their attitudes, beliefs, experiences, emotions, and actions. SA and opinion mining are two valuable methods for discerning emotions and perspectives in social media data. The insights derived from these methods can be extremely useful in various domains, such as elections, public opinion, advertising, health and treatment, and consumer satisfaction. (Sun *et al.* 2014).

Researchers in the field of SA identify the application of opinion mining and SA methods to unstructured data as a significant challenge. SA typically targets the sentence level, aspect level, document level, and user level. This can be achieved through the use of lexicon, NLP, ontology, or a combination of these and other technologies. Strategies such as feature selection, data integration, data cleansing, and crowdsourcing can be used to improve the results of SA. Figure 2.1 for a sample of SA.

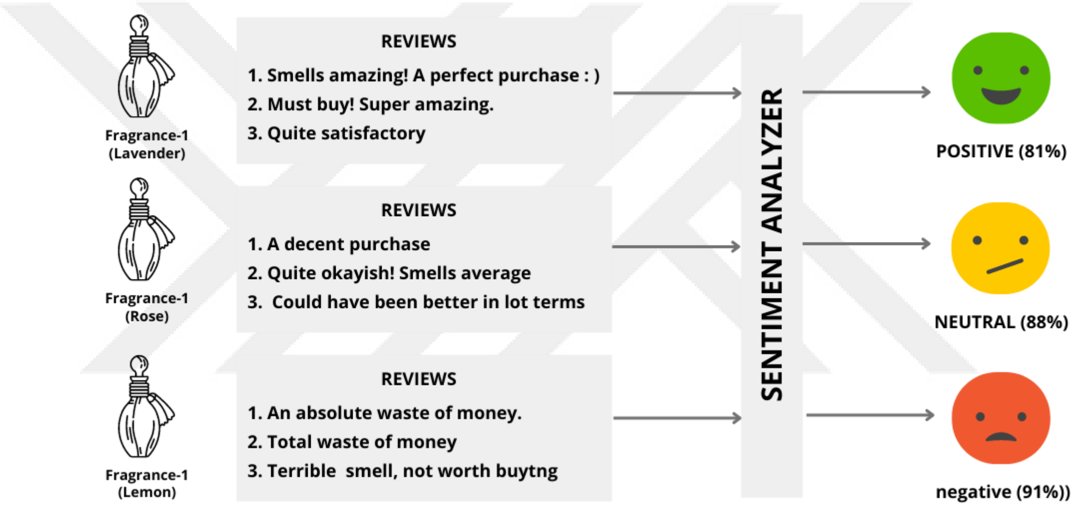


Figure 2.1 A sample of sentiment analysis

2.2.1 Sentiment analysis tasks

Emotion and polarity recognition lay the foundation for sentiment computation and analysis. The former is concerned with extracting a set of sentiment labels, whereas the latter is typically a binary classification task with outcomes such as "positive" versus "negative," "upvote" versus "downvote," or "like" versus "dislike." These two tasks are highly interconnected and dependent on one another. Some models of emotion categorization, like the 'emotion hourglass,' treat it as a single task by inferring the polarity associated with a sentence based on the emotions it communicates. In fact, emotion recognition is often considered a subtask of polarity recognition (Cambria *et al.* 2017).

Classifying polarities is an intermediate step towards conducting more complex studies. It can be used, for example, to identify "pros and cons" phrases in individual reviews to investigate the positive and negative factors affecting assessments of a product, thereby making these judgments more reliable. Another type of binary emotion classification is agreement detection, where two emotional inputs are compared to determine if they should be labeled with the same or different emotion. Complementing the classification of binary emotions, varying degrees of positivity can be assigned to the identified polarity or polarity to the inferred emotions (Cambria *et al.* 2013). If we abandon the premise that the input under evaluation is theoretical and about a single topic or instance, new complex challenges arise, such as identifying individuality, recognizing the object of opinion, and so forth. For more accurate emotion classification, it can be beneficial to ascertain whether an input is subjective or objective (Cambria *et al.* 2017). Moreover, polarity can exist in a document without the presence of an opinion; for instance, a news article can be objectively characterized as either good or bad news.

Numerous studies have demonstrated that managing these two tasks in conjunction can yield positive results. For instance, (Cambria *et al.* 2013) found that off-topic sections of documents containing irrelevant emotional content could contribute to incorrect overall sentiment polarity towards the main issue. Furthermore, a document may discuss more than one topic of interest to the reader. In such cases, it's crucial to identify the themes and segregate the sentiments accordingly. Facet extraction is a subtask of SA that, much like theme identification, seeks to pinpoint the targets of opinions in an opinionated text, such as the features of a product or service being reviewed (Cambria *et al.* 2017).

SA encompasses numerous sub-tasks such as aspect extraction, subjectivity detection, idea extraction, named entity recognition, sarcasm detection, and complementary tasks like personality detection, user profiling, and notably multimodal composition (Mohammad 2016). The widespread adoption of cameras in consumer devices like smartphones, tablets, and laptops has led to a departure from text-only communication in online social media (Cambria *et al.* 2017).

In addition to speech-to-text recognition, other forms of auditory information, facial expression and body movement analysis, or even the 'mood' conveyed by background music or color filters can be considered (Cambria *et al.* 2017). Multimodal fusion is the process of combining various signals to form a comprehensive picture. Both feature-level fusion and decision-level fusion have been explored to enhance confidence in emotion recognition from multimodal information. However, the linguistic data is typically obtained through the transcription of actual speech, rather than via automatic speech recognition.

2.2.2 Levels of sentiment analysis

Applying SA to large datasets can lead to many discoveries and economic benefits. SA, also known as opinion mining or sentiment detection, typically applies to unstructured online texts such as data from microbloggers and social media data streams. It is the process of extracting sentiments from text (Serrano-Guerrero *et al.* 2015).

Tan *et al.* (2011) presented outline four distinct contexts in which SA might be applied. The first, at the sentence level, identifies positive, negative, and neutral sentiments for individual sentences. The second, the document level, recognizes the overall tone of the document as a whole, determining whether it is positive, negative, or neutral. The third level, known as the aspect level, is applicable when features are contained within the entity, post, or input text; each feature may hold a sentiment. For instance, in a customer review of a mobile phone, features such as battery life or screen brightness can be mentioned. The sentiment attached to each feature can vary significantly. For example, while 'Nice to meet you!' and 'My phone is very interesting' are both positive at the sentence level, the overall language may need improvement at the document level. Considering the aspect level can enhance the analysis and results. Take, for example, this granular review: 'My phone is great, but it has a terrible battery. Although its apps are slow, I love its screen.' Here, the phone is the entity, and its battery life, app performance, and screen are aspects. Emotion recognition can detect negative sentiments about battery life, app performance, and positive sentiment about screen quality. At the facet level, many SA methods use clustering, where features with the same sentiment result are grouped together.

At the user level, graph theory is applied (Tan *et al.* 2011) to manage the social connections between various users. For instance, consider two individuals: User A and User B. User A often shares and likes User B's posts, and frequently references User B in their own posts. It's possible that User A and User B share similar thoughts and feelings. This might occur as a direct result of User B's influence on User A. A person in this situation can affect User B's perspective. The user-level SA takes such influence into account.

2.2.3 Process of sentiment analysis

NLP techniques are frequently used in SA to decipher and comprehend the sentiment expressed in text. There are various steps in the procedure (Hussein 2018):

- Text preprocessing involves cleaning the text data by removing unnecessary elements such as stopwords, punctuation, and other special characters.
- Tokenization is a text analysis technique that breaks down the text into its individual words or terms.
- Features are extracted from texts, which can include keywords, n-grams, and parts of speech.
- SA classifies the emotional tone of each text instance using pre-trained ML models. This can be achieved through supervised learning, where models are trained on labelled data, or by using pre-trained models that have learned sentiment patterns from large datasets.
- Additional processing can be performed on the SA data after it is collected, such as averaging sentiment scores or using threshold criteria to classify sentiments as positive, negative, or neutral.
- The effectiveness of the SA model is assessed using metrics including accuracy, precision, recall, and F1 score.

2.2.4 Types of sentiment analysis

Depending on the scope and purpose of the research, various types of SA can be conducted. Some of the most common ones include (Hussein 2018):

- Document-Level Sentiment Analysis DLSA: This form of analysis determines the overall sentiment conveyed in a review or article. The entirety of the text is analyzed to classify it as positive, negative, or neutral (Wankhade *et al.* 2022).
- Sentence-Level Sentiment Analysis SLSA: This form of analysis examines the sentiment of a whole text at the sentence level. It allows for a more nuanced understanding of the sentiments expressed in different sections of the text (Perikos and Hatzilygeroudis 2016).
- Aspect-Based Sentiment Analysis ABSA: This method identifies and extracts the sentiment associated with certain aspects or entities mentioned in the text. For instance, the evaluation of a product can be broken down into its performance, design, and usability (Perikos and Hatzilygeroudis 2017).
- Entity-Level Sentiment Analysis ELSA: This type of analysis determines the sentiment toward specific individuals, organizations, or products mentioned in the text. It's useful for understanding the range of emotions expressed in a single piece of writing (Saraff *et al.* 2018).
- Comparative Sentiment Analysis CSA: This technique compares the sentiments expressed about different characters or entities in the text. It seeks to determine how people feel about certain topics or attributes (Cambria *et al.* 2013).

2.2.5 Applications of sentiment analysis

SA has significant applications in diverse fields such as business, government, and health. Business intelligence and online commerce allow organizations to gather and analyze customer feedback to improve their services and products. SA can be used to gauge customer sentiment about an event or a product. The results of SA can provide insights into consumer preferences and opinions about market trends. Recommender systems, a subfield of AI, also utilize SA. An example of such application is survey SA using recurrent neural networks (Prithi *et al.* 2017). In broad terms, the following fields readily lend themselves to SA.

1) Social applications

Several social applications utilize SA. Current uses include monitoring student sentiment in educational fields, predicting election outcomes, recommending vacation destinations based on the satisfaction of previous visitors, and monitoring global violence through the analysis of violent tweets.

2) Medical and health applications

SA can be used on medical data to identify and predict rates of suicide and depression; tweets can be analyzed to monitor and track healthy and unhealthy locations; and doctors can be ranked based on patient satisfaction. (Satapathy *et al.* 2017) explored the use of SA in health-related social media and blog posts. The author introduced innovative approaches to text processing and ML and proposed a medical lexicon to assist professionals and individuals in navigating the extensive medical terminology currently used to describe symptoms and diseases. In the field of mental health, SA is utilized to enhance or even substitute traditional surveys by examining online posts made by patients.

3) Industrial applications

SA boasts numerous practical applications in business, encompassing brand tracking, stock market prediction, Twitter-based box office predictions, and customer satisfaction surveys. Alongside behavioral analysis, SA also offers valuable insights for the product and advertising sectors (Phan *et al.* 2020).

2.3 Sentiment Analysis Techniques

There exist various methodologies that can be employed in the conduct of SA. Figure 2.2 shows all techniques (Birjali *et al.* 2021).

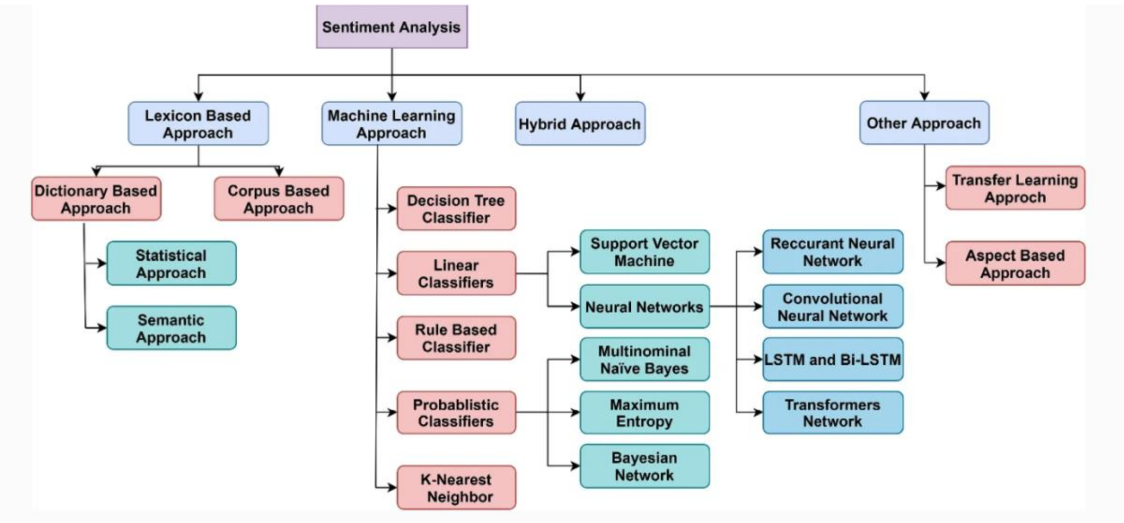


Figure 2.2 Sentiment analysis techniques (Birjali *et al.* 2021)

2.3.1 Machine learning based sentiment analysis

ML-based SA has revolutionized the industry by enabling the creation of robust and reliable models for sentiment categorization (Onan 2021). In this approach, models are trained using labeled data, where each instance is assigned a sentiment label such as "positive," "negative," or "neutral." This is achieved through the use of supervised learning techniques. Using these annotated examples, the model can generalize patterns and make predictions about the sentiment of unseen text.

Feature extraction is a vital part of supervised learning for SA. ML techniques require data to be transformed from its original textual form into a numerical one (Prema *et al.* 2021). Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) are two common feature extraction methods that describe texts as a collection of word frequencies and assess the relevance of words in a document relative to the entire corpus, respectively.

Various classification techniques have been applied in the field of SA, each with varying degrees of success. Common options include Neural Networks, Random Forests, Naive Bayes, and Support Vector Machines (SVM) (Jihad *et al.* 2023). To identify the best-performing model for a specific SA dataset, researchers often experiment with different classifiers and hyperparameters.

2.3.2 Lexicon-based approach

The polarity of phrases is determined using a set of preset dictionaries. The polarity of a document or sentence is assumed to be determined by the polarity of all phrases and words in a lexically-based approach. This method relies heavily on the meaning of words to decipher feelings. The two main categories that describe this method are the dictionary-based and the corpus-based categories. The dictionary-based method starts with the identification of the target words' roots, followed by a lookup of the dictionary for related terms. The Corpus method, on the other hand, begins with a set of foundational comments and then uses a vast corpus to locate additional comment terms in the text. The corpus-based strategy takes a statistical or semantic approach to labeling words as neutral or positive.

A limited number of seed words serve as the basis for these methods. The data in this set is then either manually or mechanically augmented. The first is a "dictionary-based" technique, whereas the second is a set-based one (Zucco *et al.* 2020). Sentiwordnet, Q-wordnet, WordNet-Affect, and SentiStrength are just few of the open-access vocabularies that may help with emotion extraction. These programs analyze the sentiment of a text by counting how often certain words (such furious, sad, amazing, and depressed) appear. Words in such a dictionary would be annotated with information such as polarity, semantic orientation score, or reliability label.

These knowledge-based methods are widely used due to their convenience and low cost. Affective terms like happy, sad, fearful, and bored are used to categorize text into distinct emotional states. Emotional vocabularies, linguistic annotation systems, and other probabilistic knowledge bases learned from linguistic corpora are often used to identify affective words or multi-word phrases (Zucco *et al.* 2020).

The biggest drawback of these methods is the lack of emotion detection when language rules are involved (Cambria *et al.* 2017). While a database may accurately label the phrase "Today was a happy day" as such, it is far less likely to do so with the phrase "Today was not a happy day at all." To determine the context of each knowledge base entry's appearance in the text, more advanced knowledge-based techniques utilize linguistic rules. Knowledge-based (dictionary) techniques heavily rely on the quality and range of their sources to be credible. In fact, it is difficult for a sentiment mining system to comprehend meaning associated with natural language or human behavior without a comprehensive knowledge base that incorporates human expertise. Another drawback of knowledge-based approaches is that the way in which information is represented is often rigidly defined, making it difficult to account for nuanced concepts and the inference of semantic and emotional features associated with them.

2.3.3 Hybrid approach

The Hybrid Approach to SA is a technique that improves upon the performance of sentiment classification by combining the advantages of ML -based approaches with lexicon-based methods (Dhaoui *et al.* 2017). By fusing the two approaches, hybrid models attempt to address the limitations inherent in each method.

Hybrid approaches often start by using ML classifiers to identify sentiments in a sentence. These classifiers can take advantage of features like word embeddings to capture context and semantic relationships between words, making them robust and accurate in many cases (Giatsoglou *et al.* 2017). However, ML models may struggle with specific sentiment modifiers, rare words, or domain-specific terms not present in their training data.

Hybrid models incorporate lexicon-based systems as a complementary step to mitigate these limitations,. The lexicon-based component can handle domain-specific terms effectively, as it relies on predefined sentiment lexicons. It can also address specific linguistic nuances that might be challenging for ML models to capture accurately.

3. MATERIALS AND METHODS

Our study process included gathering a Twitter and categorizing it using a lexicon. The proportion of positive and negative opinions was taken from the vocabulary after categorization. In this section, we have detailed the steps of the proposed method. Its general flowchart is shown in Figure 3.1.

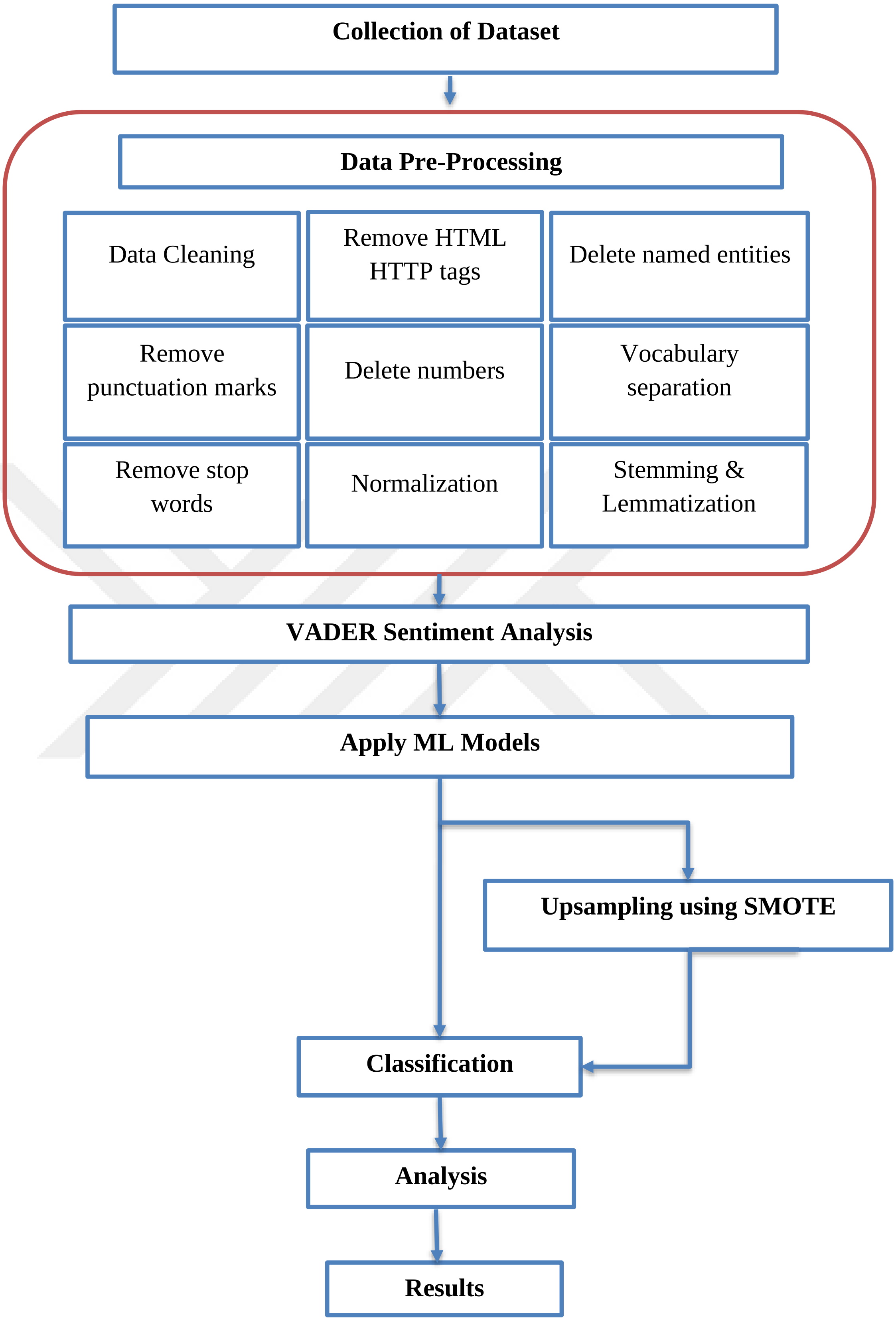


Figure 3.1 The proposed research framework

3.1 Collection of Dataset

Data collected from the Twitter API can be a valuable resource for SA and other NLP tasks. The Twitter API allows developers to access a wide variety of data related to tweets, including the text of the tweets, the user who posted the tweet, the time and date the tweet was posted, and metadata such as the number of likes, retweets, and replies.

Considering that the number of positive and neutral data in this collection is much more than the negative data, therefore one of the important challenges in this section is the imbalance of the data set, in this research, due to the existing challenges. In the data set, we use the Recall criterion to evaluate the model.

3.2 Data Pre-processing

3.2.1 Data cleaning

All social media tweet data Twitter often contains a large number of words and characters that are ineffective for data analysis (Birjali *et al.* 2021). For example, there are data tweets such as “@safemoonjustv?, hilari and educ. The data found useless words or characters such as “@” and “?”. Data cleaning techniques paired with regex can find superfluous characters and remove them from the core data to enhance the dataset’s quality.

3.2.2 Remove html http tags

Unstructured text often contains a large amount of noise, especially if techniques such as web or page scraping are used. HTML and HTTP tags are usually components that do not add much value to the understanding and analysis of the text. Therefore, in this section, we remove the unnecessary tags and keep the textual information in all the documents.

3.2.3 Delete named entities

In every text document, there are special terms that represent special institutions, and have a more informative aspect and have a unique framework. These entities are called named entities. This term specifically refers to objects in the real world such as people, places, organizations, etc., which often have specific names. Because these entities do not provide us with meaningful information for SA, so we remove them in this section.

3.2.4 Remove punctuation marks

We remove all sets of punctuation marks including [~{|}'_[\]@?<=>;:/.-,+*()'&%$# "!] Because these signs are seen in all sentences, removing them will lead to positive results.

3.2.5 Delete numbers

During text preprocessing, we remove the numbers in the text data that are not related to the text analysis and do not lead to the production of meaningful information.

3.2.6 Vocabulary separation

Vocabulary separation, also referred to as Tokenization, involves breaking down larger blocks of text into smaller, more manageable units of language called tokens. Tokens represent the smallest linguistic units and can be anything from words and numbers to punctuation marks.

3.2.7 Remove stop words

Stop words are actually words that are commonly used. Words that are meaningless or have no special meaning, especially when semantic features are extracted from the text. These items are usually very frequent in the text, and usually these words include adjectives, conjunctions, additions and such. Some examples of stop words include and, the, an, a, and the like. During NLP, there is no tendency for these types of words to occupy space or take up valuable processing time. For this reason, these words can be easily removed in this section.

Prepositions, conjunctions, adjectives, slank words, pronouns, and many more words are quite useful. These terms are typically seen combined with the primary word; therefore, it is not distinctive and has no special meaning. A stop word or stop list is a list of words that do not contribute much to analytical content.

3.2.8 Normalization

In this subsection, all the letters in the text data are converted to lowercase letters.

3.2.9 Stemming and lemmatization

In every language, words will have different appearances according to the role they play in sentences. But considering that all of them are made from the same root, they will help us in terms of meaning and concept. Therefore, in many NLP-based methods, we must first find the root of the words. The act of rooting makes it possible to convert different forms of a word into a single form. With this, the number of features is reduced and also the different forms of a word are removed and the computer can consider different forms of a word as one. The process of returning words to their root form is called the operation of lexical rooting and semantic rooting, so these two methods are usually used to find the root of words. There are different algorithms to perform lexical rooting. Porter’s algorithm is very famous in English. According to a series of regular rules, this algorithm can obtain the roots of words with good accuracy. Semantic rooting can also be done by methods. In this practice, it is necessary to use a dictionary or something similar to obtain the roots of words, because generally, the methods of finding semantic roots are not regular.

- Lemmatization: This is a process in NLP where words are reduced to their base or root form, called 'lemma.' It helps in standardizing words to their canonical form, which is linguistically correct. For example, the words "running," "runs," and "ran" would all be converted to their base form "run." This process is crucial in SA, as it allows for the grouping and analysis of similar sentiments expressed through variations of a word. While it doesn't directly handle slang or non-standard terminology, it helps to unify different forms of standard words, thereby enhancing the accuracy of the SA.

- Stemming: Stemming is a process where words are shortened by removing their prefixes or suffixes. This process aims to reduce a word to its stem or root form, which may not necessarily be a valid word on its own. For instance, stemming could reduce the word influences to the simpler form "influence". Search engines and other text analysis tools often use stemming to improve the efficiency and relevancy of their results.

3.3 Select the Important Features

The TF-IDF (term frequency-inverse document frequency) vectorizer was employed in an application. TF-IDF is a metric that quantifies the significance of a word within the context of a document in a corpus. It is achieved by multiplying the term frequency (TF) a measure of how often a word occurs in a document by the inverse document frequency (IDF), which gauges how uncommon a term is across the corpus. This computation yields the document term weight, not density (DT). A collection of text documents can be transformed into a matrix of TF-IDF features using the TF-IDF vectorizer, with each row representing a document and each column a word. A SA ML model can take this matrix as input and provide an interpretation of the data. In SA and other NLP applications, the TF-IDF vectorizer is a widely used feature extractor (Rajput *et al.* 2021).

3.4 VADER Sentiment Analysis

VADER, or Valence Aware Dictionary and Sentiment Reasoner, is a lexicon and rule-based SA tool that performed exceptionally well on social media SA according to (Alamoodi *et al.* 2022). A distinctive feature of VADER is that it eschews polarity from the document scoring process, instead providing positive, negative, neutral, and compound scores. An advantage of VADER is that it doesn't require training data, thus enabling us to apply it to previously unseen data.

This system is capable of detecting both the polarity (positive/negative) and intensity (strength) of emotion. It is included in the NLTK (Natural Language Toolkit) package and can be applied to unlabeled text data without any preprocessing. To enhance the accuracy of SA, they assign values to each word based on whether it is positive, negative, or neutral. For instance, on a numeric scale, the positive word 'good' was assigned an emotional weight of 0.52. Adding an intensifier such as 'so' increased the score by 0.61, compared to when only the word 'good' was used. Conversely, the word 'bad' was given a negative score of -0.48, as it is typically associated with negative sentiments. This scoring approach better reflects the emotional weight each word carries. VADER categorized tweets about AI assistants into positive, negative, and neutral groups, and assessed the sentiment of each document matrix. If a tweet contained no positive or negative terms, the matrix would register a score of 'zero'. Additionally, the system could provide a percentage indicating how many of the words used were positive, negative, or neutral.

3.5 ML Models

We investigate several popular ML algorithms for SA in this article. SA, a form of opinion mining, utilizes NLP to identify the positive, negative, or neutral emotions expressed in a piece of text. We will delve into the functionalities of four widely-used ML models for SA Like (Logistic regression, XGBoost, AdaBoost and SGD).

3.5.1 Logistic regression

Logistic regression (LR) is a simple linear classification model that makes predictions based on a weighted combination of input features (Hosmer *et al.* 2013). For SA, LR takes in a text input and predicts whether the sentiment is positive or negative based on the presence and frequency of certain words and phrases (Medhat *et al.* 2014). It is easy to implement and interpret, but struggles to handle complex linguistic nuances. LR has been widely applied as a baseline sentiment classifier (Araque *et al.* 2017). However, it is generally outperformed by more advanced methods.

3.5.2 XGBoost

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance (Chen and Guestrin 2016). For SA, XGBoost builds an ensemble of decision trees where each tree learns from the errors of the previous one to make increasingly accurate predictions. XGBoost performs well on SA tasks across many domains and datasets (Ke *et al.* 2017). It captures non-linear relationships and interactions between words better than linear models like LR. XGBoost is also efficient to train and tune, making it popular for sentiment modeling.

3.5.3 AdaBoost

AdaBoost, short for Adaptive Boosting, is another ensemble method that combines multiple weak learners, typically decision trees (Freund and Schapire, 1997). For SA, AdaBoost iteratively trains classifiers on modified versions of the data that emphasize previously misclassified examples (Esuli and Sebastiani, 2009). This focus on hard examples improves accuracy. AdaBoost has outperformed LR and Naive Bayes models for sentiment classification (Bangyal *et al.* 2021).

3.5.4 SGD

SGD optimizes models by taking small steps along estimated gradients, leading to fast model updates (Bottou 2010). For SA, SGD enables efficient training of linear classifiers like LR on large datasets (Umer *et al.* 2021). It is easy to implement and appropriate for text data where examples are high-dimensional and sparse. However, convergence can be slow and tuning is required to control step sizes. SGD remains a popular optimization technique but is often used within more advanced neural network architectures.

3.6 SMOTE

SMOTE (Synthetic Minority Oversampling Technique) is a popular resampling method used to balance class distributions in datasets with a skewed class imbalance (Chawla *et al.* 2002).

SMOTE addresses class imbalance by generating new synthetic minority class samples rather than merely duplicating existing samples as in basic oversampling. The algorithm selects a minority class sample and computes its k-nearest neighbors from the minority class. New samples are interpolated between the selected sample and its neighbors.

3.7 Evaluation Criteria

Various criteria have been introduced to evaluate the performance of the text classification system.

The accuracy criterion expresses the number of correct predictions made by the classifier, divided by the total number of predictions made by the same classifier. In general, accuracy refers to how well the model predicts the output, Equation (3.1) represent to accurecy calcultion (Baker *et al.* 2022).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{3.1}$$

Where TP is ture possitive, TN is true negative, FP is false positive and FN is false negative.

As shown in Equation (3.2), the Recall criterion is the ratio of the number of text data correctly classified in a specific class to the total number of data that should be classified in the same specific class. When the False Negatives value is high, the Recall criterion will be a suitable criterion (Nigam *et al.* 2000).

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{3.2}$$

The measure F1 criterion combines accuracy and correctness parameters to determine how well a classification model performs. Compared to the accuracy criterion, measure F1 draws a more accurate picture of how the classification model works on all the classes in the data. The F1 criterion is one at best and zero at worst, Equation (3.3) shows the calaulation of F1-score (Adak *et al.* 2022).

$$\text{F1} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \tag{3.3}$$

When the classes in the data are unbalanced, the accuracy criterion is not a suitable criterion to evaluate the performance of a classifier. Therefore, to evaluate the performance of our proposed model, considering that the data set used is unbalanced, we use the average accuracy criterion.

4. RESULTS AND DISCUSSION

We present the findings of our thesis in this section. As mentioned previously, we employed VADER SA to assign labels to our cleaned dataset. This dataset was subsequently utilized in ML models, both with and without the application of the Synthetic Minority Over-sampling Technique (SMOTE) for balancing. Lastly, we define the evaluation metrics we used to gauge the accuracy and efficiency of these models.

4.1 Dataset labeling and using VADER

In this subsection, we describe the process of labeling our dataset, which is a crucial step in supervised ML tasks. The dataset was initially raw and unannotated, and we needed to assign appropriate labels to each instance to indicate the target variable or outcome we are trying to predict.

The labeling process involved human annotators who were provided with clear guidelines and context to assign labels based on the specific task. Depending on the nature of the study, the labeling could be binary (e.g., positive/negative, yes/no) or multiclass (e.g., low/medium/high).

To ensure consistency and accuracy in the labeling process, we performed regular quality checks and resolved any discrepancies through discussions with the annotators. The final labeled dataset became the foundation for our supervised ML models.

In addition to the target variable labels, we also used VADER for SA on certain text data within our dataset. SA helps us understand the emotional tone or polarity of the text, which can be valuable in various applications, such as customer reviews, social media SA, and market research.

Using VADER, we assigned a sentiment score to each text instance in the dataset. The scores ranged from -1 (most negative) to +1 (most positive), with 0 indicating a neutral sentiment. VADER utilizes a lexicon and rule-based approach, which makes it particularly suitable for social media text and short informal texts.

This SA using VADER allowed us to explore the emotional trends within the data and potentially find correlations between sentiment and the target variable (if applicable). It provided additional insights that could be used in combination with other features for more robust predictive modeling. Figure 4.1 shows the results of applying VADER on our dataset. The figure shows that the majority of tweets are positive, on the other hand the minority of tweets are negative.

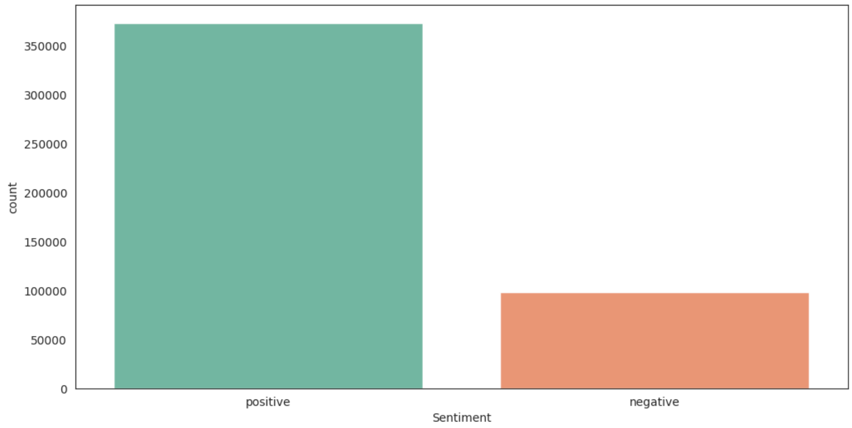


Figure 4.1 VADER sentiment analysis

Figure 4.2 shows a word cloud visualization of words frequently associated with negative sentiment tweets. The dominant terms include "ref", "bad", "worst", "horrible", "terrible", "devil" and "pain" which indicate strong negative emotions. On Other terms like "money", "cost" and "price" suggest dissatisfaction regarding financial elements. The prevalence of words like "not", "don't", "can't" and "wouldn't" imply negation and refusal. Overall, the word cloud provides a summarized snapshot of key terms in tweets expressing negative opinions and attitudes.

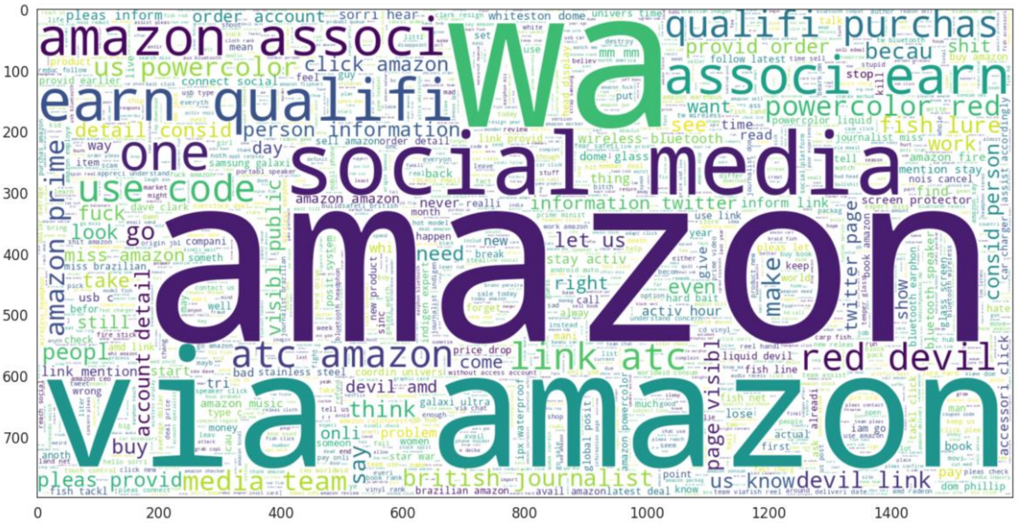


Figure 4.2 Word cloud sentiment in negative

Figure 4.3 shows a wordcloud for terms associated with positive sentiment. Dominant words include "love", "like", "good", "best", "great", "happy", "amazing" which convey strong positive emotions. The terms "thank", "help" and "support" imply gratitude and appreciation. Words such as "friend", "family" and "child" suggest positive affiliations and relationships. The prevalence of present tense verbs indicates optimism and enthusiasm. Overall this word cloud summarizes key terms in tweets with favorable and upbeat sentiments.

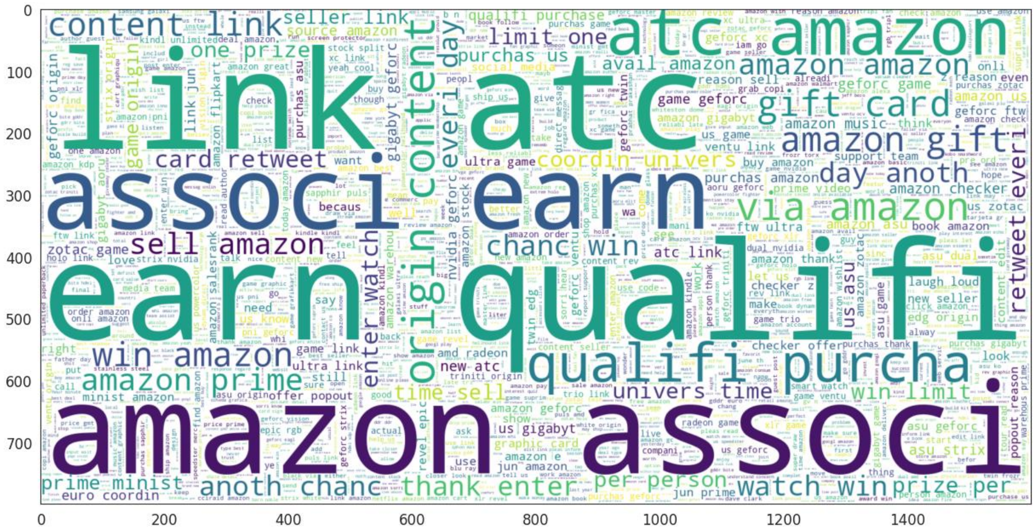


Figure 4.3 Word cloud sentiment in positive

Comparing the two graphs, the negative sentiment terms appear more diffuse while the positive words are more concentrated around a few main concepts like "love", "like" and "good". The positive terms also seem more personal and relationship-oriented while the negative words relate more to general emotions. The word clouds provide an efficient visualization of the lexical differences between tweets of opposing sentiment. Analyzing the frequent terms provides useful insights into how sentiment is expressed in the text data.

4.2 Applying ML Models on the Dataset without Using SMOTE

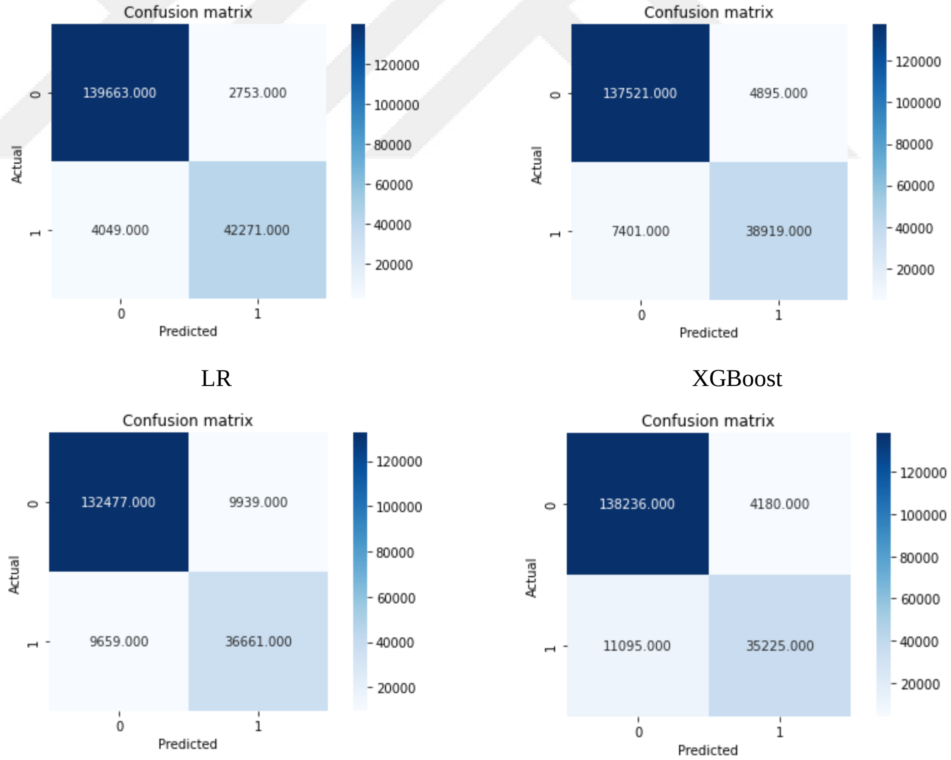
In this section, we tested several ML models on the raw data without adjusting the distribution of the variables. The aim was to evaluate the performance of these models despite the imbalanced class composition. We employed commonly used classifiers such as LR, XGBoost, AdaBoost, and SGD. Evaluation metrics included accuracy, precision, recall, F1-score, cross-validation, and the area under the Receiver Operating Characteristic (ROC) curve. These metrics are outlined in Table 4.1.

Table 4.1 Result analysis of sentiment analysis by splitting dataset to 60 Training 40 Testing (without using SMOTE)

PROPOSED ALGORITHMS	ACCURECY	PRECISION	RECALL	F1-SCORE	MCC	AUC
LR	0.963	0.963	0.963	0.963	0.901	0.947
XGBoost	0.934	0.934	0.934	0.934	0.821	0.903
AdaBoost	0.896	0.896	0.896	0.896	0.720	0.861
SGD	0.919	0.917	0.919	0.919	0.774	0.866

From Table 4.1, we observe that LR achieved the highest accuracy score of 0.963. This indicates that LR correctly classified 96.3% of the instances in the dataset. XGBoost and SGD also demonstrated strong accuracy, with values of 0.934 and 0.919, respectively. However, AdaBoost showed a slightly lower accuracy of 0.896. In addition, we can see that all algorithms achieved high precision scores, indicating their ability to minimize false positives. LR, XGBoost, and AdaBoost performed equally well with precision values of 0.963, while SGD obtained a slightly lower precision of 0.917. Looking at the results, we find that LR, XGBoost, and SGD achieved the highest recall scores of 0.919.

This suggests that these algorithms were effective at capturing most of the positive instances. AdaBoost, on the other hand, obtained a recall score of 0.896, which was slightly lower. From Table 4.1, we see that LR, XGBoost, and SGD obtained identical F1-scores of 0.919, indicating a balance between precision and recall. AdaBoost, with an F1-score of 0.896, exhibited slightly lower performance in this aspect. The Matthews Correlation Coefficient (MCC) takes into account both true and false positives and negatives, making it a valuable metric for assessing model performance, especially in imbalanced datasets. Based on the results, LR achieved the highest MCC of 0.901, followed by XGBoost (0.821), SGD (0.774), and AdaBoost (0.720). LR achieved the highest AUC of 0.947, suggesting good overall discrimination. XGBoost and SGD also demonstrated competitive AUC values of 0.903 and 0.866, respectively. However, AdaBoost had a relatively lower AUC of 0.861. Figure 4.4 shows the Confusion matrix for applied ML models.



LR

XGBoost

AdaBoost

SGD

Figure 4.4 Confusion matrix for applied ML models

Figure 4.5 shows the ROC plot results for obtained ML models after applying 60:40 before applying SMOTE. A higher score signifies a classifier's increased capability to differentiate between the two classes being measured. Therefore, based on this criteria, LR was determined to have classified the positive class of the dataset more accurately.

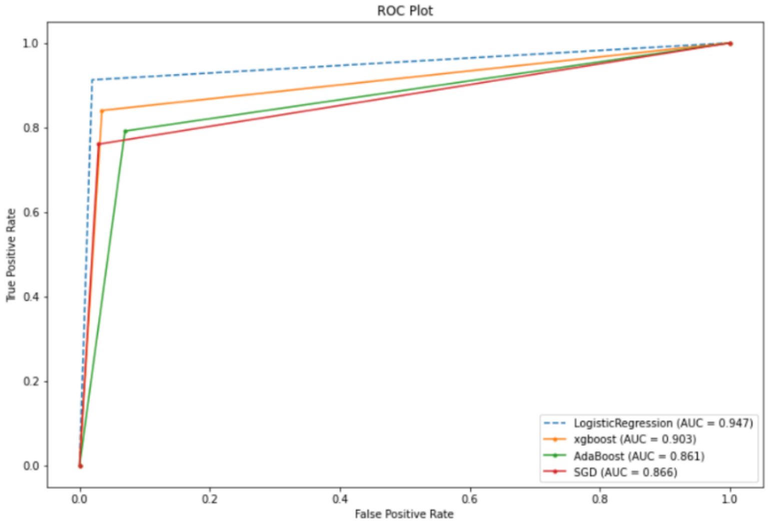


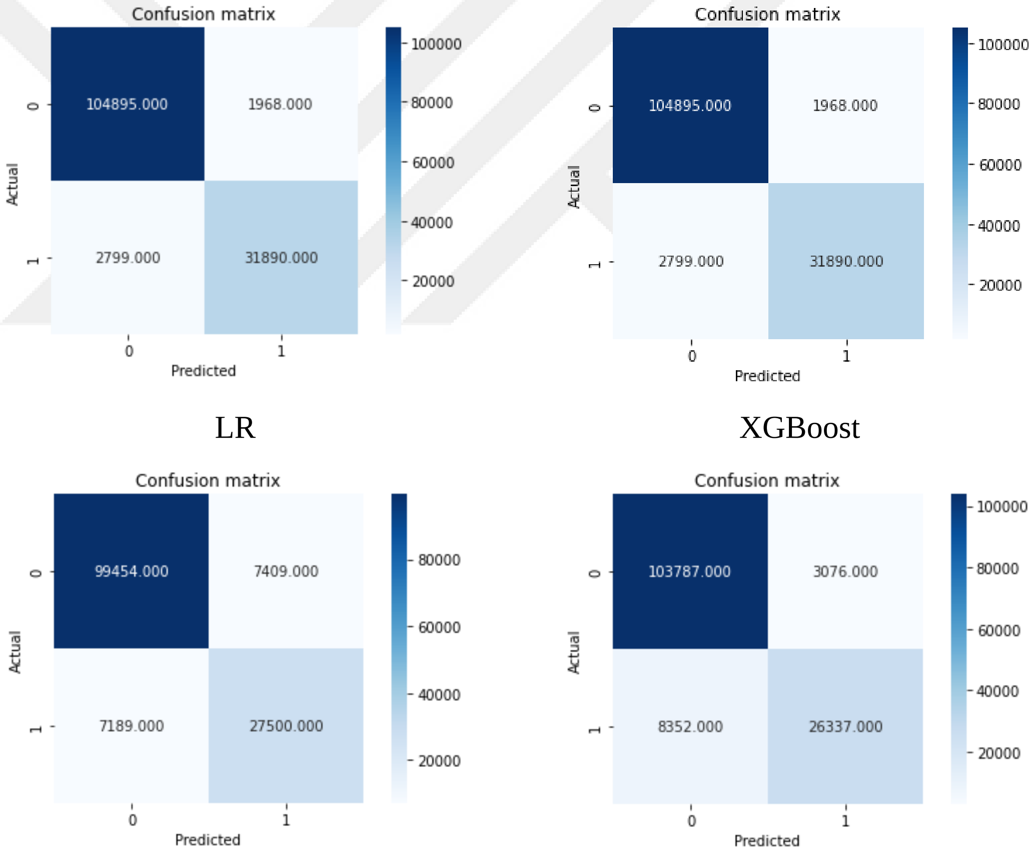
Figure 4.5 The ROC plot results for obtained ML models after applying 60:40 before applying SMOTE

In addition, we split our dataset into 70:30 to check the performance of used ML models. Table 4.2 shows the result analysis of SA by splitting dataset to 70 Training 30 Testing (without using SMOTE).

Table 4.2 Result analysis of sentiment analysis by splitting dataset to 70 Training 30 Testing (without using SMOTE)

PROPOSED ALGORITHMS	**ACCURECY**	**PRECISION**	**RECALL**	**F1-SCORE**	**MCC**	**AUC**
LR	0.966	0.966	0.966	0.966	0.908	0.950
XGBoost	0.938	0.937	0.938	0.938	0.830	0.907
AdaBoost	0.896	0.897	0.896	0.896	0.721	0.862
SGD	0.919	0.918	0.919	0.919	0.774	0.865

The accuracy metric represents the overall correctness of the model's predictions. From the results of table 4.2, we observe that LR achieved the highest accuracy of 0.966, followed by XGBoost with an accuracy of 0.938. SGD obtained an accuracy of 0.919, while AdaBoost had the lowest accuracy of 0.896. LR, XGBoost, and AdaBoost demonstrated similar precision scores of around 0.966 and 0.897, respectively. SGD achieved a slightly lower precision of 0.918. LR, XGBoost, and SGD achieved recall scores of approximately 0.966 and 0.919, respectively, while AdaBoost obtained a recall score of 0.896. All algorithms exhibited high F1-scores, with LR, XGBoost, and SGD obtaining scores of around 0.966 and AdaBoost achieving a score of 0.896. LR achieved the highest MCC of 0.908, followed by XGBoost (0.830), SGD (0.774), and AdaBoost (0.721). ın addition, LR achieved the highest AUC of 0.950, suggesting excellent discrimination. XGBoost and SGD also demonstrated competitive AUC values of 0.907 and 0.865, respectively, while AdaBoost had a relatively lower AUC of 0.862. Figure 4.6 shows the Confusion matrix for applied ML models.



LR XGBoost

Adaboost SGD

Figure 4.6 Confusion matrix for applied ML models

Figure 4.7 shows the ROC plot results for obtained ML models after applying 70:30 before applying SMOTE. A higher score signifies a classifier's increased capability to differentiate between the two classes being measured. Therefore, based on this criteria, LR was determined to have classified the positive class of the dataset more accurately.

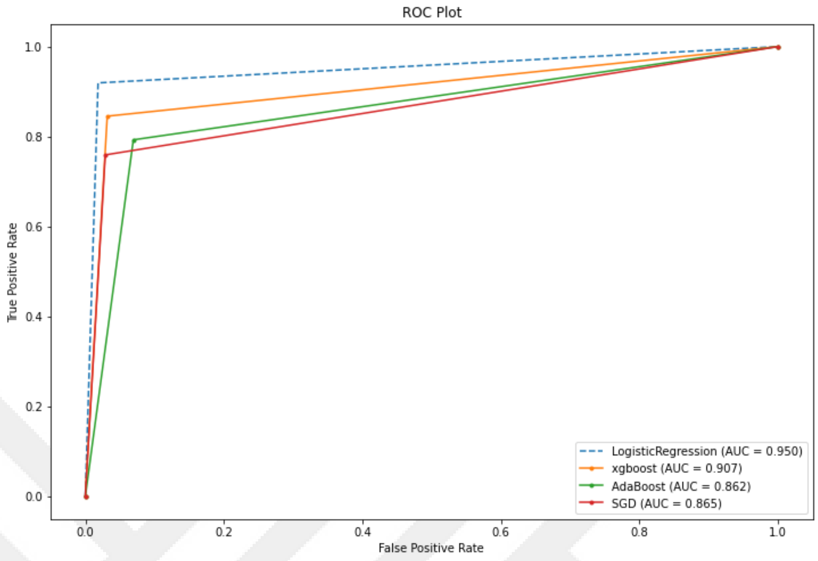


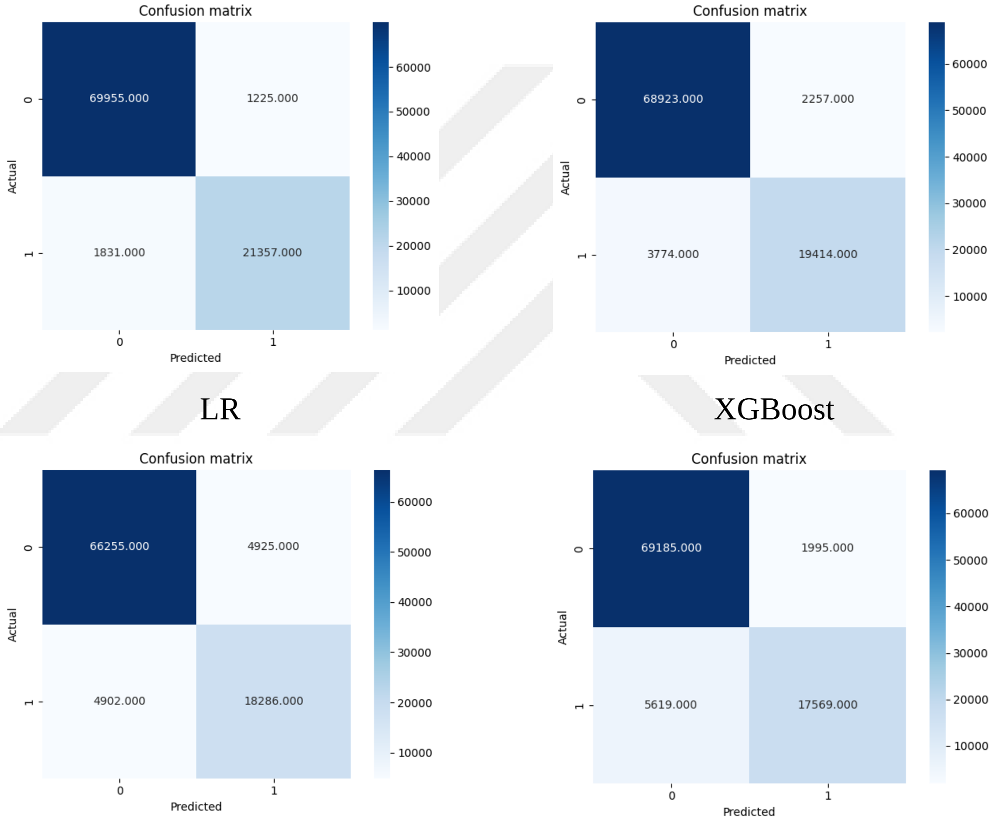
Figure 4.7 The ROC plot results for obtained ML models after applying 70:30 before applying SMOTE

Next we split our dataset into 80:20 to check the performance of used ML models. Table 4.3 shows the result analysis of SA by splitting dataset to 80 Training 20 Testing (Without using SMOTE).

Table 4.3 Result analysis of sentiment analysis by splitting dataset to 80 Training 20 testing (without using SMOTE)

PROPOSED ALGORITHMS	ACCURECY	PRECISION	RECALL	F1-SCORE	MCC	AUC
LR	0.960	0.967	0.967	0.967	0.911	0.952
XGBoost	0.936	0.935	0.936	0.936	0.824	0.903
AdaBoost	0.895	0.895	0.895	0.895	0.719	0.860
SGD	0.919	0.918	0.919	0.919	0.774	0.865

It can conduct from Table 4.3 that LR achieved an accuracy of 0.960, making it the highest among the four algorithms. XGBoost had an accuracy of 0.936, SGD achieved 0.919, and AdaBoost had the lowest accuracy of 0.895. LR, XGBoost, and AdaBoost demonstrated similar precision scores of around 0.967 and 0.895, respectively, while SGD achieved a slightly lower precision of 0.918. LR, XGBoost, and SGD achieved recall scores of approximately 0.967 and 0.919, respectively, while AdaBoost obtained a recall score of 0.895. In addition, All algorithms exhibited high F1-scores, with LR, XGBoost, and SGD obtaining scores of around 0.967 and AdaBoost achieving a score of 0.895. LR achieved the highest MCC of 0.911, followed by XGBoost (0.824), SGD (0.774), and AdaBoost (0.719). In terms of AUC, LR achieved the highest AUC of 0.952, suggesting excellent discrimination. XGBoost and SGD also demonstrated competitive AUC values of 0.903 and 0.865, respectively, while AdaBoost had a relatively lower AUC of 0.860. Figure 4.8 shows the Confusion matrix for applied ML models.



LR XGBoost

Adaboost SGD

Figure 4.8 Confusion matrix for applied ML models

Figure 4.9 shows ROC plot the results for obtained ML models after applying 80:20 before applying SMOTE. The higher the value for a classifier, the better its ability to distinguish between positive and negative classes. So it can be said that LR has done a better job in classifying the positive class in the dataset.

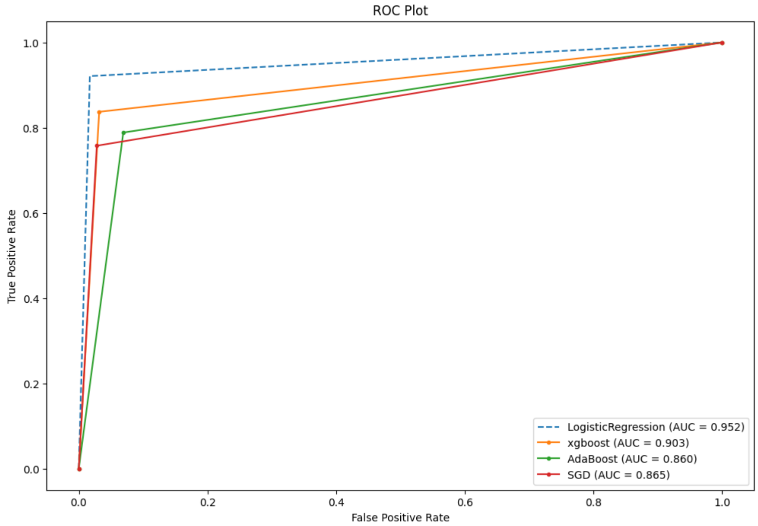


Figure 4.9 The ROC plot results for obtained ML models after applying 80:20 before applying SMOTE

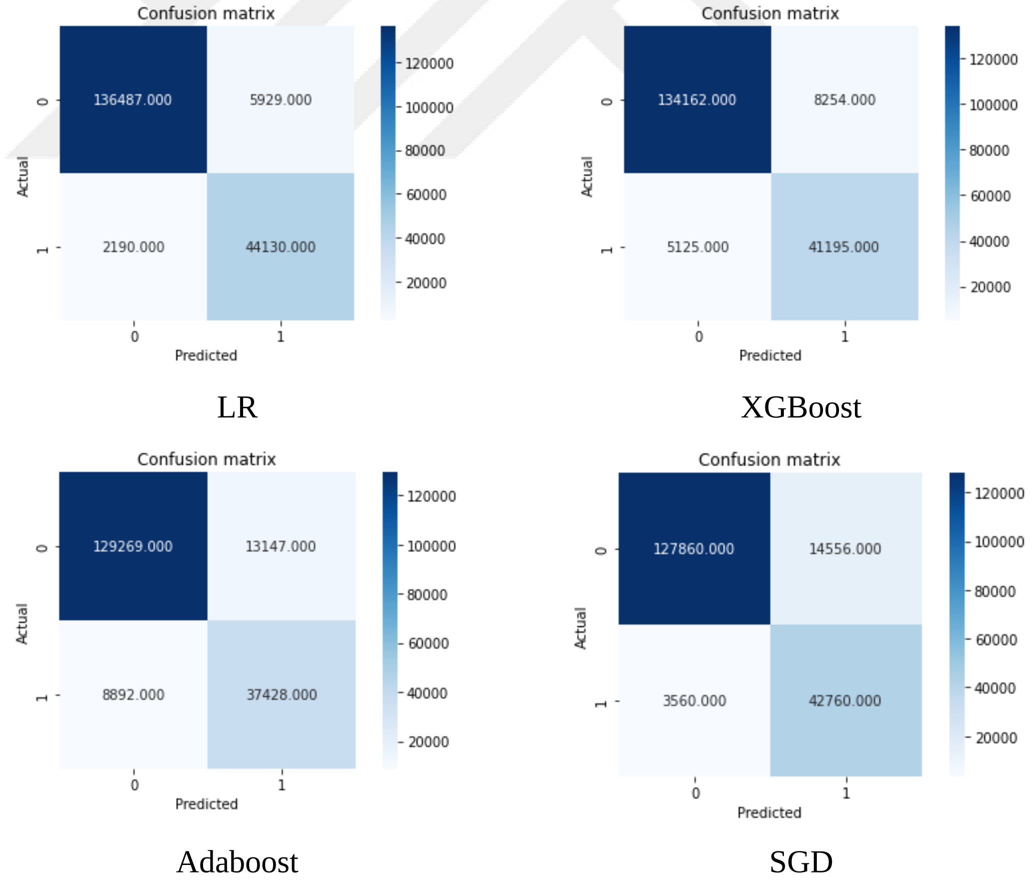
4.3 Applying ML models on the dataset using SMOTE

We utilized the Synthetic Minority Over-sampling Technique (SMOTE) to rectify the class imbalance in our dataset. SMOTE creates artificial samples for the underrepresented group by extrapolating between nearby real-world instances. This section explains the implementation of SMOTE and its impact on student enrollment. Subsequently, we applied the same family of ML models to the SMOTE-balanced dataset. The objective of this study was to ascertain whether balancing the dataset could improve model performance. We present the final results, including metrics such as accuracy, precision, recall, F1-score, Matthew's Correlation Coefficient (MCC), and Area under the ROC Curve (AUC), in Table 4.4.

Table 4.4 Result analysis of sentiment analysis by splitting dataset to 60 Training 40 testing (using SMOTE)

PROPOSED ALGORITHMS	**ACCURECY**	**PRECISION**	**RECALL**	**F1-SCORE**	**MCC**	**AUC**
LR	0.956	0.959	0.956	0.956	0.888	0.956
XGBoost	0.929	0.931	0.929	0.929	0.813	0.916
AdaBoost	0.883	0.887	0.883	0.883	0.695	0.858
SGD	0.904	0.917	0.904	0.904	0.768	0.910

From Table 4.4, we observe that LR achieved an accuracy of 0.956, making it the highest among the four algorithms. XGBoost had an accuracy of 0.929, SGD achieved 0.904, and AdaBoost had the lowest accuracy of 0.883. LR, XGBoost, and AdaBoost demonstrated similar precision scores of around 0.959 and 0.887, respectively, while SGD achieved a slightly higher precision of 0.917. LR, XGBoost, and SGD achieved recall scores of approximately 0.956 and AdaBoost obtained a recall score of 0.883. All algorithms exhibited high F1-scores, with LR, XGBoost, and SGD obtaining scores of around 0.956 and AdaBoost achieving a score of 0.883. LR achieved the highest MCC of 0.888, followed by XGBoost (0.813), SGD (0.768), and AdaBoost (0.695). LR achieved the highest AUC of 0.956, suggesting excellent discrimination. XGBoost and SGD also demonstrated competitive AUC values of 0.916 and 0.910, respectively, while AdaBoost had a relatively lower AUC of 0.858. Figure 4.10 shows the Confusion matrix for applied ML models.



LR

XGBoost

Adaboost

SGD

Figure 4.10 Confusion matrix for applied ML models

Figure 4.11 shows the ROC plot results for obtained ML models after applying 60:40 after applying SMOTE. If a classifier has a higher value, it is better equipped to differentiate between positive and negative categories. Consequently, we can conclude that LR effectively classifies the positive class of the dataset.

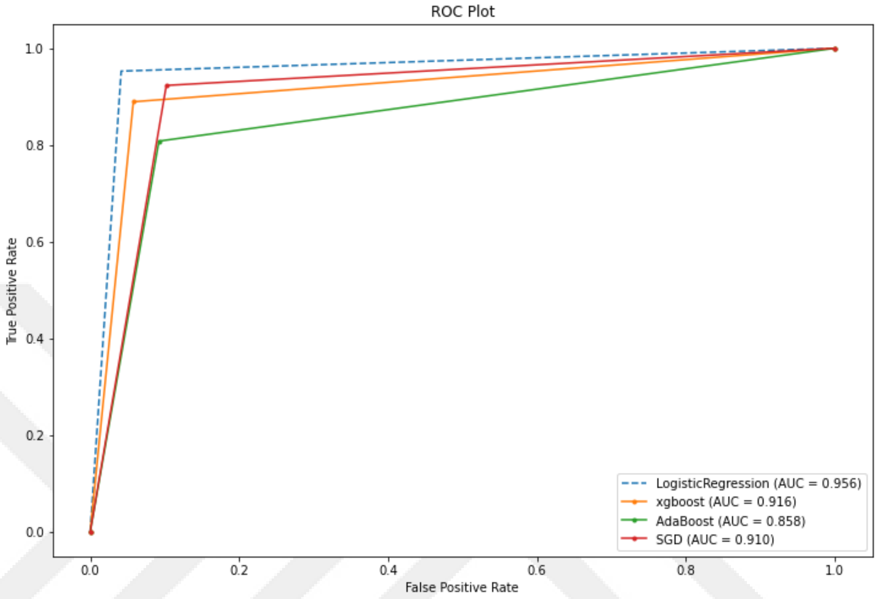


Figure 4.11 The ROC plot results for obtained ML models after applying 60:40 after applying SMOTE

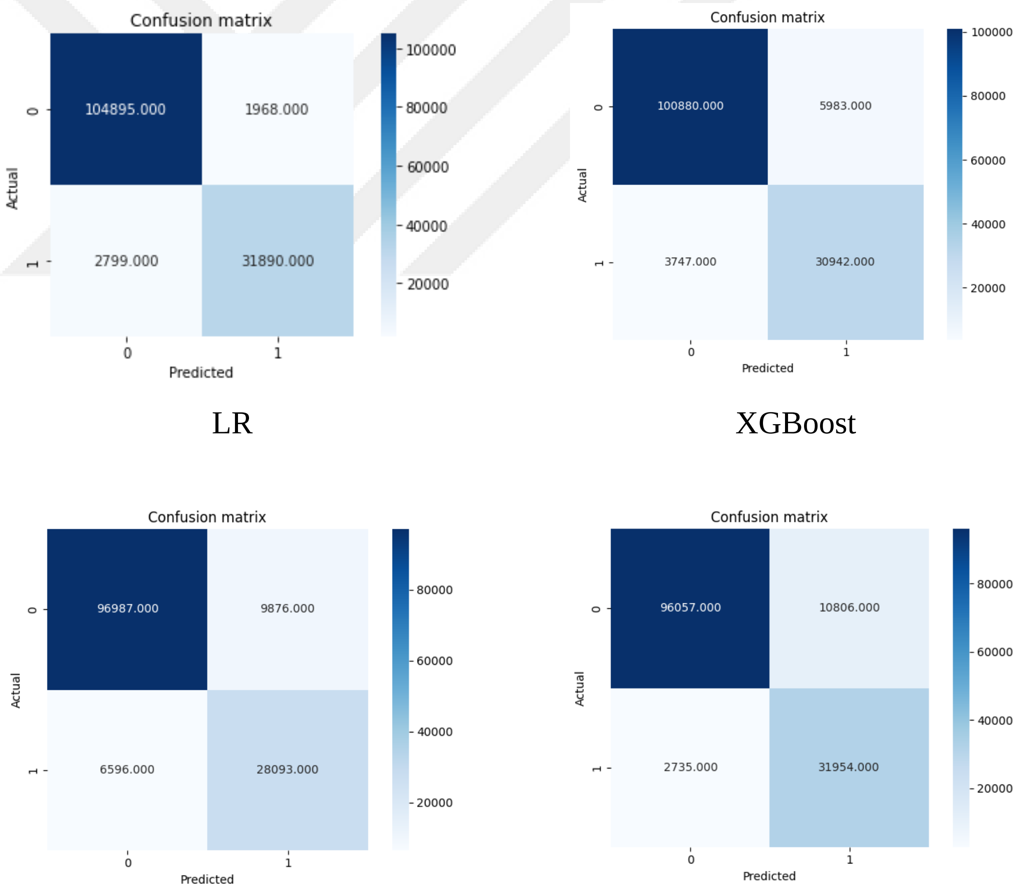
In addition, we split our dataset into 70:30 to check the performance of used ML models. Table 4.5 shows the result analysis of SA by splitting dataset to 70 Training 30 Testing (using SMOTE).

Table 4.5 Result analysis of sentiment analysis by splitting dataset to 70 training 30 testing (without using SMOTE)

PROPOSED ALGORITHMS	ACCURECY	PRECISION	RECALL	F1-SCORE	MCC	AUC
LR	0.959	0.961	0.959	0.959	0.893	0.957
EGB	0.931	0.933	0.931	0.931	0.818	0.918
Adaboost	0.883	0.888	0.883	0.883	0.696	0.859
SGD	0.904	0.917	0.904	0.904	0.768	0.910

From the results of Table 4.5, we observe that LR achieved an accuracy of 0.959, making it the highest among the four algorithms. EGB had an accuracy of 0.931, SGD achieved

0.904, and AdaBoost had the lowest accuracy of 0.883. LR, EGB, and AdaBoost demonstrated similar precision scores of around 0.961 and 0.888, respectively, while SGD achieved a slightly higher precision of 0.917. LR, EGB, and SGD achieved recall scores of approximately 0.959 and AdaBoost obtained a recall score of 0.883. All algorithms exhibited high F1-scores, with LR, EGB, and SGD obtaining scores of around 0.959 and AdaBoost achieving a score of 0.883. LR achieved the highest MCC of 0.893, followed by EGB (0.818), SGD (0.768), and AdaBoost (0.696). LR achieved the highest AUC of 0.957, suggesting excellent discrimination. EGB and SGD also demonstrated competitive AUC values of 0.918 and 0.910, respectively, while AdaBoost had a relatively lower AUC of 0.859. Figure 4.12 shows the Confusion matrix for applied ML models.



LR

XGBoost

Adaboost

SGD

Figure 4.12 Confusion matrix for applied ML models

Figure 4.13 shows the ROC plot results for obtained ML models after applying 70:30 after applying SMOTE. The higher the value for a classifier, the better its ability to distinguish between positive and negative classes. So it can be said that LR has done a better job in classifying the positive class in the dataset.

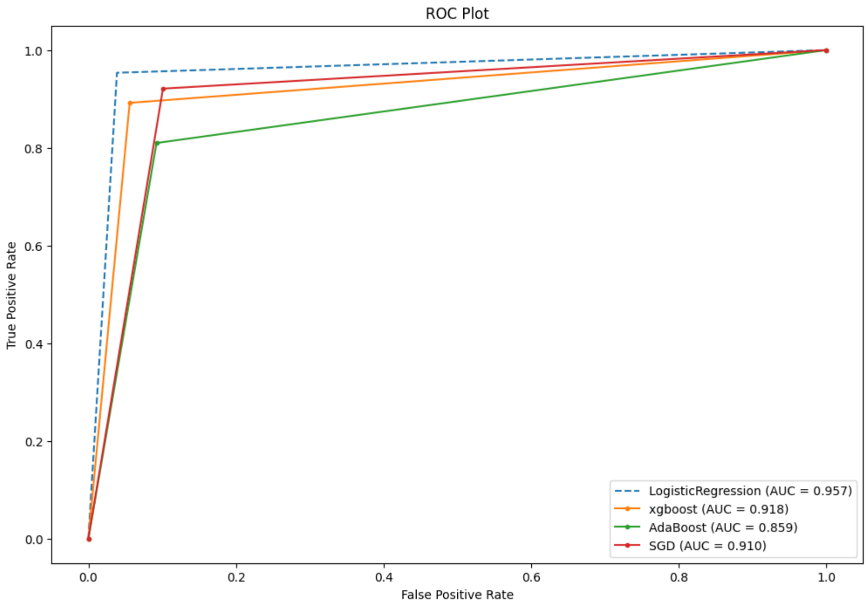


Figure 4.13 The ROC plot results for obtained ML models after applying 70:30 after applying SMOTE

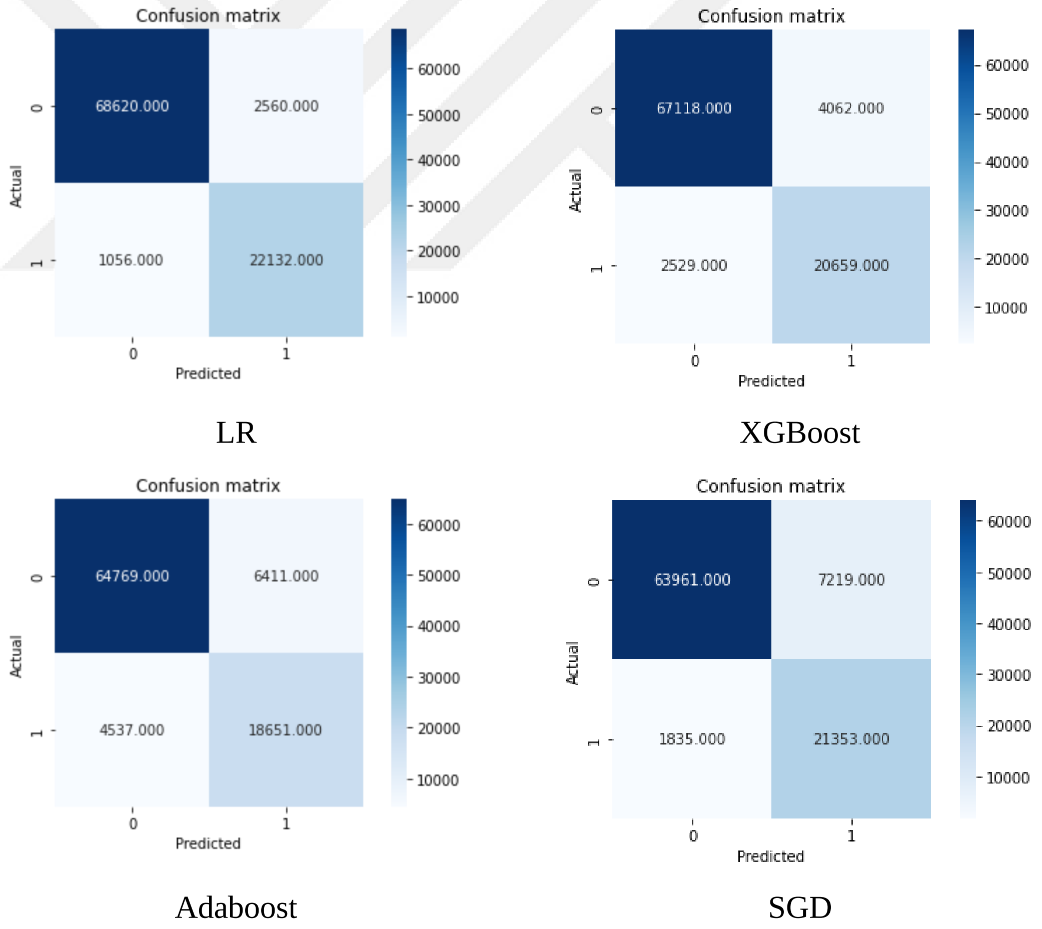
Next, we split our dataset into 80:20 to check the performance of used ML models. Table 4.6 shows the result analysis of SA by splitting dataset to 80 Training 20 Testing (using SMOTE).

Table 4.6 Result analysis of sentiment analysis by splitting dataset to 80 training 20 testing (using SMOTE)

PROPOSED ALGORITHMS	ACCURECY	PRECISION	RECALL	F1-SCORE	MCC	AUC
LR	0.961	0.963	0.961	0.961	0.899	0.959
XGBoost	0.930	0.932	0.930	0.930	0.816	0.917
Adaboost	0.883	0.887	0.883	0.883	0.696	0.857
SGD	0.904	0.916	0.904	0.904	0.767	0.910

It can conduct from Table 4.6 that LR achieved an accuracy of 0.961, making it the highest among the four algorithms. XGBoost had an accuracy of 0.930, SGD achieved

0.904, and AdaBoost had the lowest accuracy of 0.883. LR, XGBoost, and AdaBoost demonstrated similar precision scores of around 0.963 and 0.887, respectively, while SGD achieved a slightly higher precision of 0.916. LR, XGBoost, and SGD achieved recall scores of approximately 0.961 and AdaBoost obtained a recall score of 0.883. All algorithms exhibited high F1-scores, with LR, XGBoost, and SGD obtaining scores of around 0.961 and AdaBoost achieving a score of 0.883. LR achieved the highest MCC of 0.899, followed by XGBoost (0.816), SGD (0.767), and AdaBoost (0.696). the ROC Curve (AUC): LR achieved the highest AUC of 0.959, suggesting excellent discrimination. XGBoost and SGD also demonstrated competitive AUC values of 0.917 and 0.910, respectively, while AdaBoost had a relatively lower AUC of 0.857. Figure 4.14 shows the Confusion matrix for applied ML models.



LR

XGBoost

Adaboost

SGD

Figure 4.14 Confusion matrix for applied ML models

Figure 4.15 shows the ROC plot results for obtained ML models after applying 80:20 after applying SMOTE. A higher score indicates that a classifier is more capable of differentiating between the two classes being measured. Therefore, LR has been assessed to classify the positive class of the dataset more accurately.

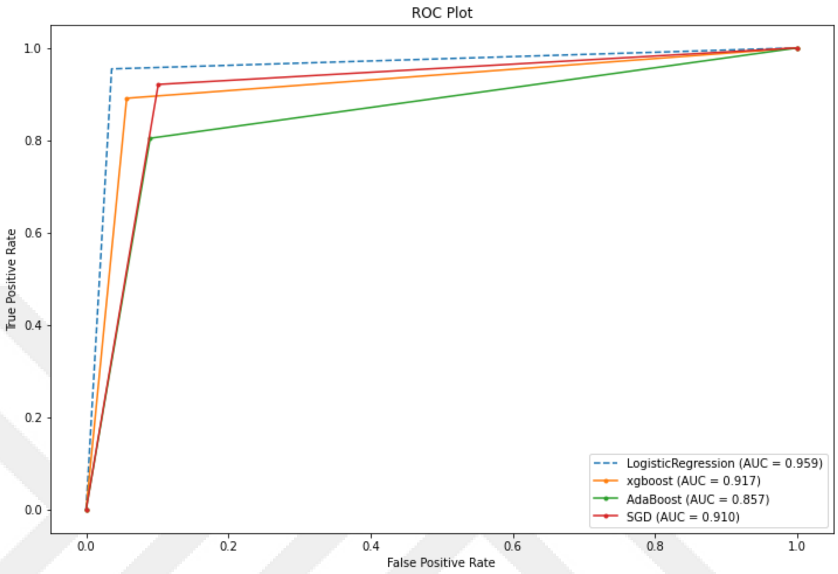


Figure 4.15 The ROC plot results for obtained ML models after applying 80:20 after applying SMOTE

From our experiments, it is evident that LR consistently outperformed the other algorithms across most evaluation metrics. LR demonstrated the highest accuracy, precision, recall, F1-score, MCC, and AUC, making it the most suitable choice for this particular task. XGBoost and SGD also exhibited competitive performance and can be considered viable alternatives, especially when interpretability is less critical and higher model complexity is acceptable.

However, it is important to note that the performance of the algorithms may vary depending on the specific characteristics of the dataset and the nature of the task. Therefore, it is recommended to conduct further experiments on different datasets and tasks to validate the generalizability of the results.

5. CONCLUSIONS AND RECOMMENDATION

In this work, we performed an automatic tagging and SA method on raw Twitter data related to e-commerce activity. This was done using VADER for sentiment polarity detection, along with four ML algorithms. Before data analysis, the data was cleaned through processes like folding, data deletion, rewording, stop word removal, and stemming. Once cleaned, the data could be automatically tagged and classified with the four ML algorithms. The combined workflow of VADER sentiment polarity detection and ML algorithms effectively analyzed raw Twitter data across three scenarios. From our experiments, it is evident that LR consistently outperformed the other algorithms across most evaluation metrics. It demonstrated the highest accuracy, precision, recall, F1-score, MCC, and AUC, making it the optimal choice for this particular task. XGBoost and Stochastic Gradient Descent (SGD) also exhibited competitive performance, and could be viable alternatives, particularly when interpretability is less critical and a higher degree of model complexity is acceptable. However, it's important to note that the performance of these algorithms may vary depending on the specific characteristics of the dataset and the nature of the task. Therefore, we recommend conducting further experiments on different datasets and tasks to validate the generalizability of the results.

We can summarize the results as following:

- LR achieved the best results overall based on the evaluation metrics. It had the highest accuracy, precision, recall, F1-score, MCC and AUC.
- XGBoost and SGD also performed reasonably well on this dataset.
- AdaBoost lagged behind the other algorithms in terms of performance.
- Additional experiments on different data are needed to fully validate the findings and determine if they generalize across tasks.
- Dataset characteristics and problem type can impact relative model performance.
- So while LR appears optimal for this case, the other algorithms may prove superior given different criteria and data. More research is required to fully compare the models' applicability.

In the future, we will strive to apply various ways to make our model more efficient, as well as employ other aspect-based analytic methodologies on negative review data sets to identify problematic product attributes. As a result, the business may improve the quality of its products and boost its sales rate.

The techniques and algorithms used for SA are rapidly advancing, but there are still many unsolved problems in the field of opinion research. Based on this, in this part and according to the axes of this research, titles with the aim of continuing research and developing knowledge in the field of research; It is suggested to the next researchers.

1) Visualizing the results obtained in this research with the aim of helping users to better understand the feelings extracted from the text of the comments, on the target website can be an effective step in researching opinions about products and services.
2) The results obtained in relation to the analysis of sentiments around products and services are very important, and organizations, companies and e-commerce websites, as well as research in the field of digital marketing, can benefit from these results and in line with Take more effective steps to promote digital marketing.
3) It is still difficult to identify fake comments from other comments, and the effect of the positive or negative polarity of these comments on the final result is uncontrollable. Therefore, the identification of fake opinions in the field of analysis of Turkish language should be considered as a very important field of study by researchers.
4) Every opinion text with a positive polarity can be considered as a valuable resource, which always suggests products or services to other users and customers. Based on this, companies, organizations, as well as e-commerce websites can benefit from this comment text and with the help of recommender systems, products and services to other customers and users who choose the category. Products and services have the same taste as the people providing the opinion.
5) Creating and developing tools with high efficiency and accuracy to convert colloquial expressions into official expressions and a tool to detect spelling mistakes in the text of comments.

6) It is suggested that other researchers should also use the word cloud technique in addition to the opinion analysis steps they are considering for their research, so that the audience can understand the content mentioned in relation to the text of the comments in a more comprehensible way.

7) E-commerce websites can take effective steps in attracting more and better users and customers by using the word cloud technique and using the text of the comments on the website, because the word cloud technique, in addition to having visual appeal, helps users and It helps customers to understand in a shorter time the features mentioned in the comments about products and services.

REFERENCES

Adak, A., Pradhan, B., Shukla, N. and Alamri, A. 2022. Unboxing deep learning model of food delivery service reviews using explainable artificial intelligence (XAI) technique. Foods, 11(14).

Alamoodi, A. H., Baker, M. R., Albahri, O. S., Zaidan, B. B. and Zaidan, A. A. 2022. Public sentiment analysis and topic modeling regarding COVID-19's three waves of total lockdown: A case study on movement control order in malaysia. KSII Transactions on Internet and Information Systems, 16(7): 2169–2190.

Alamoodi, A., Zaidan, B., Zaidan, A., Albahri, O., Mohammed, K., Malik, R., Almahdi, E., Chyad, M., Tareq, Z. and Albahri, A. 2020. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. Expert Systems with Applications, 114155.

Alasiri, M. M. Salameh, A. A. 2020. The impact of business intelligence (BI) and decision support systems (DSS): Exploratory study. International Journal of Management, 11(5): 1001–1016.

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F. and Iglesias, C. A. 2017. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 77: 236–246.

Baker, M. R., Mahmood, Z. N. and Shaker, E. H. 2022. Ensemble Learning with Supervised Machine Learning Models to Predict Credit Card Fraud Transactions. Revue d'Intelligence Artificielle, 36(4): 509–518.

Baker, M. R., Mohammed, E. Z. and Jihad, K. H. 2023. Prediction of Colon Cancer Related Tweets Using Deep Learning Models. In Intelligent Systems Design and Applications. ISDA 2022. Lecture Notes in Networks and Systems (pp. 522–532). Springer, Cham.

Bangyal, W. H., Qasim, R., Rehman, N. U., Ahmad, Z., Dar, H., Rukhsar, L., Aman, Z. And Ahmad, J. 2021. Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches. Computational and Mathematical Methods in Medicine, 2021.

Berisha-Shaqiri, A. and Berisha-Namani, M. 2015. Information Technology and the Digital Economy. Mediterranean Journal of Social Sciences, 6: 2039–2117.

Bottou, L. 2010. Large-scale machine learning with stochastic gradient descent. Proceedings of COMPSTAT 2010 - 19th International Conference on Computational Statistics, Keynote, Invited and Contributed Papers, 177–186.

Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A. 2017. Affective computing and sentiment Analysis (pp. 1–10).

Cambria, E., Poria, S., Hazarika, D. and Kwok, K. 2018. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. 32nd AAAI Conference on Artificial Intelligence, AAAI 2018: 1795–1802.

Cambria, E., Schuller, B., Xia, Y. and Havasi, C. 2013. New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2): 15–21.

Can, U. and Alatas, B. 2019. A new direction in social network analysis: Online social network analysis problems and applications. Physica A: Statistical Mechanics and Its Applications, 535.

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. 2002. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16: 321–357.

Chen, T. and Guestrin, C. 2016. XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-Augu, 785–794.

Cirqueira, D., Hofer, M., Nedbal, D., Helfert, M. and Bezbradica, M. 2020. Customer purchase behavior prediction in e-commerce: A conceptual framework and research agenda. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11948: 119–136.

Dhaoui, C., Webster, C. M. and Tan, L. P. 2017. Social media sentiment analysis: lexicon versus machine learning. Journal of Consumer Marketing, 34(6): 480–488.

Dong, Y. and Jiang, W. 2019. Brand purchase prediction based on time-evolving user behaviors in e-commerce. Concurrency and Computation: Practice and Experience, 31(1).

Esuli, A. and Sebastiani, F. 2009. Training data cleaning for text classification. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5766: 29–41.

Fan, W. and Gordon, M. D. 2014. The power of social media analytics. Communications of the ACM, 57(6): 74–81.

Fisher, N. I. and Kordupleski, R. E. 2019. Good and bad market research: A critical review of Net Promoter Score. Applied Stochastic Models in Business and Industry, 35(1), 138–151.

Freund, Y. and Schapire, R. E. 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1): 119–139.

Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G. and Chatzisavvas, K. C. 2017. Sentiment analysis leveraging emotions and word embeddings. Expert Systems with Applications, 69: 214–224.

Gunasekaran, A., Marri, H. B., McGaughey, R. E. and Nebhwani, M. D. 2002. E-commerce and its impact on operations management. International Journal of Production Economics, 75(1–2): 185–197.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. 2013. Applied logistic regression: third edition. In Applied Logistic Regression: Third Edition. wiley.

Hovy, D. and Yang, D. 2021. The Importance of Modeling Social Factors of Language: Theory and Practice. NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 588–602.

Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 168–177.

Jihad, K. H., Baker, M. R., Farhat, M. and Frikha, M. 2023. Machine learning-based social media texta: Impact of the rising fuel prices on electric vehicles. In Hybrid Intelligent Systems. HIS 2022. Lecture Notes in Networks and Systems pp. 625–635. Springer, Cham.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T. Y. 2017. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 2017: 3147–3155.

Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S. and Khan, K. H. 2016. Sentiment analysis and the complex natural language. Complex Adaptive Systems Modeling, 4(1).

Kumar, A. and Sebastian, T. M. 2012. Sentiment Analysis: A Perspective on its Past, Present and Future. International Journal of Intelligent Systems and Applications, 4(10): 1–14.

Liao, S. H., Widowati, R. and Hsieh, Y. C. 2021. Investigating online social media users' behaviors for social commerce recommendations. Technology in Society, 66.

Liu, B. 2012. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1): 1–184.

London, T. and Hart, S. L. 2004. Reinventing strategies for emerging markets: Beyond the transnational model. Journal of International Business Studies, 35(5), 350–370.

Medhat, W., Hassan, A. and Korashy, H. 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4): 1093–1113.

Mehta, P., Pandya, S. and Kotecha, K. 2021. Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Computer Science, 7: 1–21.

Micol Policarpo, L., da Silveira, D. E., da Rosa Righi, R., Antunes Stoffel, R., da Costa, C. A., Victória Barbosa, J. L., Scorsatto, R. and Arcot, T. 2021. Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. Computer Science Review, 41.

Mohammad, S. M. 2016. Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. In Emotion Measurement pp. 201–237.

Nigam, K., Mccallum, A. K., Thrun, S. and Mitchell, T. 2000. Text classification from labeled and unlabeled documents using EM. Machine Learning, 39(2): 103–134.

Nisar, T. M. and Prabhakar, G. 2017. What factors determine e-satisfaction and consumer spending in e-commerce retailing? Journal of Retailing and Consumer Services, 39: 135–144.

O'Cass, A. and Fenech, T. 2003. Web retailing adoption: Exploring the nature of internet users Web retailing behaviour. Journal of Retailing and Consumer Services, 10(2): 81–94.

Onan, A. 2021. Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. Computer Applications in Engineering Education, 29(3): 572–589.

Pang, B., Lee, L. and Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. 79–86.

Perikos, I. and Hatzilygeroudis, I. 2016. Recognizing emotions in text using ensemble of classifiers. Engineering Applications of Artificial Intelligence, 51: 191–201.

Perikos, I. and Hatzilygeroudis, I. 2017. Aspect based sentiment analysis in social media with classifier ensembles. Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017: 273–278.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I. and Manandhar, S. 2014. SemEval-2014 Task 4: Aspect based sentiment analysis. In: 8th International Workshop on Semantic Evaluation, SemEval 2014 - Co-Located with the 25th International Conference on Computational Linguistics, 27–35.

Prema Arokia Mary, G., Hema, M. S., Maheshprabhu, R. and Nageswara Guptha, M. 2021. Sentimental analysis of twitter data using machine learning algorithms. 2021 International Conference on Forensics, Analytics, Big Data, Security, 2021.

Rabjohn, N., Cheung, C. M. K. and Lee, M. K. O. 2008. Examining the perceived credibility of online opinions: Information adoption in the online environment. Proceedings of the Annual Hawaii International Conference on System Sciences.

Rajput, G. K., Kundu, S. and Kumar, A. 2021. The impact of feature extraction on multi-source sentiment analysis, In: Proceedings of the 2021 10th International Conference on System Modeling and Advancement in Research Trends, 2021: 510–515.

Saraff, S., Taraban, R., Rishipal, R., Biswal, R., Kedas, S. and Gupta, S. 2018. Application of Sentiment Analysis in Understanding Human Emotions and Behaviour. EAI Endorsed Transactions on Smart Cities, 166547.

Sin, S. S., Nor, K. M. and Al-Agaga, A. M. 2012. Factors affecting malaysian young consumers’ online purchase intention in social media websites. Procedia - Social and Behavioral Sciences, 40: 326–333.

Smith, D. N. and Sivakumar, K. 2004. Flow and Internet shopping behavior. A conceptual model and research propositions. Journal of Business Research, 57(10): 1199–1208.

Umer, M., Ashraf, I., Mehmood, A., Kumari, S., Ullah, S. and Sang Choi, G. 2021. Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model. Computational Intelligence, 37(1): 409–434.

Wankhade, M., Rao, A. C. S. and Kulkarni, C. 2022. A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7): 5731–5780.

Zeng, Y. E., Joseph Wen, H. and Yen, D. C. 2003. Customer relationship management (CRM) in business-to-business (B2B) e-commerce. Information Management and Computer Security, 11(1): 39–44.

Zhang, L., Wang, S. and Liu, B. 2018. Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4).

APPENDICES

APPENDIX 1. Study software and hardware information

APPENDIX 1. Software and hardware information

1) Software requirements

- Python with suitable JDK i.e., Jupyter notebook software.

2) Hardware requirements

- Graphics: NVIDIA GeForce RTX2070 8G Max-Q GDDR6 w/ New Ray Tracing Technology
- Processor: Intel Core i7-8750H 2. 2 - 4. 1GHz
- Memory: 32GB (16G*2) DDR4 2666MHz 2 Sockets; Max Memory 32GB.