THE DEVELOPMENT OF A SENTIMENT ANALYSIS MODEL FOR TEACHING AND LEARNING USING TEXT MESSAGING BASED ON THE BERT MODEL

MD ABDUL BAKIR

SULTAN IDRIS EDUCATION UNIVERSITY

2026

THE DEVELOPMENT OF A SENTIMENT ANALYSIS MODEL FOR TEACHING AND LEARNING USING TEXT MESSAGING BASED ON THE BERT MODEL

MD ABDUL BAKIR

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

DISSERTATION PRESENTED TO QUALIFY FOR A MASTER’S IN SCIENCE

(RESEARCH MODE)

FACULTY OF COMPUTING AND META-TECHNOLOGY
SULTAN IDRIS EDUCATION UNIVERSITY

2026

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

UPSI/IPS-3/BO 32
Pind : 00 m/s: 1/1

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

Please tick (√)

Project Paper	
Masters by Research	√
Master by Mixed Mode	
PhD	

INSTITUTE OF GRADUATE STUDIES

DECLARATION OF ORIGINAL WORK

This declaration is made on the ....21........day of...January..20.25...

i. Student's Declaration:

I, BAKIR MD ABDUL, M20212002953, FACULTY OF COMPUTING AND META TECHNOLOGY (PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled A DEVELOPMENT OF A SENTIMENT ANALYSIS MODEL FOR TEACHING AND LEARNING USING TEXT MESSAGING BASED ON THE BERT MODEL is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Bakir

Signature of the student

ii. Supervisor's Declaration:

I ASSOCIATE PROFESSOR DR. SULIANA BINTI SULAIMAN (SUPERVISOR'S NAME) hereby certifies that the work entitled A DEVELOPMENT OF A SENTIMENT ANALYSIS MODEL FOR TEACHING AND LEARNING USING TEXT MESSAGING BASED ON THE BERT MODEL (TITLE) was prepared by the above named student, and was submitted to the Institute of Graduate Studies as a * partial/full fulfillment for the conferment of MASTERS IN SCIENCE (SOFTWARE ENGINEERING) (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work.

21.01.2025

Date

Signature of the Supervisor

UPSI/IPS-03/BO31
Pind:02 m/s: 1/1

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

INSTITUT PENGAJIAN SISWAZAH / *INSTITUTE OF GRADUATE STUDIES*

BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM

Tajuk / *Title*: THE DEVELOPMENT OF A SENTIMENT ANALYSIS MODEL FOR TEACHING AND LEARNING USING TEXT MESSAGING BASED ON THE BERT MODEL

No. Matrik /: *Matric's No.*: M20212002953

Saya / *I am*: MD ABDUL BAKIR

(Nama Pelajar / *Student's Name*)

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana) ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut: -
acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows: -

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik Universiti Pendidikan Sultan Idris.
   The Thesis/Dissertation/Project Paper is the property of Universiti Pendidikan Sultan Idris.
2. Perpustakaan Tuanku Bainun dibenarkan membuat Salinan untuk tujuan rujukan dan pendidikan.
   Tuanku Bainun Library has the right to make copies for the purpose of reference and education.
3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi/Laporan Kertas Projek ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
   The library has the right to make copies of the thesis/Dissertation/Project Paper for academic exchange.
4. Sila tandakan (√) bagi pilihan kategori di bawah / *Please tick (√) for category below: -*

√	**AKSES TERBUKA** ***OPEN ACCESS***	Saya bersetuju bahawa Tesis/Disertasi/Kertas Projek boleh diterbitkan sebagai salinan fizikal atau akses terbuka dalam talian. *I agree that the Thesis/Dissertation/Project Paper may be published as a physical copy or open access online.*
	TERHAD** ****LIMITED	*Mengandungi maklumat terhad seperti yang dinyatakan oleh organisasi/institusi di mana penyelidikan dijalankan. (Perpustakaan akan menyekat akses sehingga ____ tahun). Contains restricted information as specified by the organization/institution where the research was conducted. (Library will restrict access until ____ years).*
	SULIT** ****CONFIDENTIAL	*Mengandungi maklumat sulit di bawah Akta Rahsia Rasmi 1972. Contains confidential information under the Official Secrets Act 1972.*
	EMBARGO	Tesis/Disertasi/Kertas Projek ini tertakluk kepada tempoh embargo selama 2 tahun atas sebab-sebab berikut. *This Thesis/Dissertation/Project Paper is subject to an embargo period of 2 years for the following reasons:* ____________

*Nota / *Notes:
i. Jika tiada pilihan yang ditandai, pilihan pertama akan dilaksanakan secara automatik.
   If no option is marked, the first option will be automatically implemented.
ii. **Jika Tesis/Disertasi/Kertas Projek ini dikategorikan sebagai SULIT atau TERHAD, sila lampirkan surat dari pihak berkuasa/organisasi yang berkaitan dengan menyatakan alasan dan tempoh tesis ini perlu diklasifikasikan sebagai TERHAD atau SULIT. / *If the Thesis/Dissertation/Project Paper is categorized as CONFIDENTAL or LIMITED, kindly attach a letter from the relevant authority/organization stating the reason and duration for the thesis should be classified as LIMITED or CONFIDENTIAL.***

(Tandatangan Pelajar/ Signature)

(Tandatangan Penyelia / Signature of Supervisor)
& (Nama & Cop Rasmi / Name & Official Stamp)

Tarikh/Date: 21.01.2025

Tarikh/Date: 22.01.2025

ACKNOWLEDGMENT

All praise is due to ALLAH (SWT), the Most Gracious, the Most Merciful, for granting me the strength, knowledge, and perseverance to complete this research. This work would not have been possible without His infinite blessings and guidance. I also express my profound appreciation for my family (especially my father, Md Rafiqul Islam, and my mother, Beauty Begum), who stepped in with emotional reassurance and critical financial support when unexpected personal challenges arose. Their unconditional love and sacrifice have been the foundation of my resilience and success. I extend my deepest gratitude to my supervisor, Associate Professor Dr. Suliana Binti Sulaiman, for her unwavering support and insightful critiques throughout my research journey. Her dedication to academic excellence and meticulous attention to detail have significantly shaped this dissertation. I am also profoundly thankful for the opportunity she provided by selecting me as a Graduate Research Assistant (GRA). I am equally grateful to my co-supervisor, Dr. Salem Abdullah Salem Garfan, for his constructive feedback and essential suggestions. I would also like to express my sincere appreciation to Mr. Sasigaran Moneyam for his continuous support and willingness to help whenever needed. His cooperation and encouragement played an essential role in facilitating various aspects. Finally, I extend my gratitude to all staff and my friends who provided academic support, constructive discussions, and moral encouragement throughout my postgraduate journey. Their contributions, whether direct or indirect, are sincerely appreciated.

ABSTRACT

This study explores sentiment analysis in educational contexts, focusing on unravelling the inherent in online learning environments. It aims to develop and evaluate a sentiment analysis model for classifying students' sentiments using data from text messaging platforms such as WhatsApp and Telegram. The challenge of sentiment analysis in educational contexts is twofold, which includes data-related issues and the effectiveness of analytical models. The primary concern is data-related issues, encompassing both the availability and the nature of the data. Educational text data are often unstructured, multilingual, and linguistically diverse, with frequent use of textisms, emojis, and mixed language expressions. These characteristics, combined with the scarcity of domain-specific datasets and the complexity of languages like Malay and English mixed language, hinder accurate sentiment detection. The research adopts systematic literature review methods, including pre-processing and developing the Bidirectional Encoder Representations from Transformers (BERT) model. This consists of pre-processing the text messenger data, fine-tuning sentiment analysis models, and evaluating the model. The results indicate that the proposed sentiment analysis with the BERT model dramatically enhances the accuracy rate in classifying up to 89% of the sentiments. The comparative analysis model of the BERT based sentiment classifier with other models, including Naive Bayes (84%), Support Vector Machine (83%), Random Forest (84%), and K-Nearest Neighbors (81%). Using this model educators can used it assist to analyse students' sentiments from WhatsApp and Telegram message to support teaching and learning activities. However, future accountability may focus on fine-grained emotion detection, larger multilingual datasets, and multimodal analytics to enhance sentiment interpretation in educational environments.

PEMBANGUNAN MODEL ANALISIS SENTIMEN UNTUK PENGAJARAN DAN PEMBELAJARAN MENGGUNAKAN MESEJ TEKS BERDASARKAN MODEL BERT

ABSTRAK

Kajian ini meneroka analisis sentimen dalam konteks pendidikan dengan memberi tumpuan kepada pengungkapan sentimen yang wujud dalam persekitaran pembelajaran dalam talian. Ianya bertujuan untuk membangun dan menilai model analisis sentimen untuk mengklasifikasikan sentimen pelajar menggunakan data daripada platform pesanan teks seperti WhatsApp dan Telegram. Cabaran analisis sentimen dalam konteks pendidikan bersifat dua hala, iaitu melibatkan isu berkaitan data dan keberkesanan model analisis. Isu utama adalah berkaitan data, yang merangkumi aspek ketersediaan dan sifat data itu sendiri. Data teks pendidikan lazimnya tidak berstruktur, bersifat pelbagai bahasa, serta mempunyai kepelbagaian linguistik, termasuk penggunaan singkatan tidak formal, emoji, dan ungkapan bahasa campuran. Bagi mengatasi permasalahan ini, kajian ini mengaplikasikan teknik pra-pemprosesan dan membangunkan model *Bidirectional Encoder Representations from Transformers* (BERT). Pendekatan ini melibatkan pra-pemprosesan data pemesejan teks, penalaan halus model analisis sentimen, serta penilaian prestasi model berkenaan. Hasil kajian menunjukkan bahawa model analisis sentimen berasaskan BERT yang dicadangkan berjaya meningkatkan kadar ketepatan pengelasan sehingga 89%. Analisis perbandingan turut dijalankan antara pengelas sentimen berasaskan BERT dengan model lain, iaitu Naive Bayes (84%), Support Vector Machine (83%), Random Forest (84%), dan K-Nearest Neighbors (81%). Dengan menggunakan model ini, pendidik boleh menggunakannya untuk membantu menganalisis sentimen pelajar melalui pemesejan WhatsApp dan Telegram untuk menyokong aktiviti pengajaran dan pembelajaran. Walau bagaimanapun, kajian akan datang boleh memfokuskan kepada pengesanan emosi secara terperinci, penggunaan set data pelbagai bahasa yang lebih besar, serta analitik multimodal bagi meningkatkan keupayaan tafsiran sentimen dalam persekitaran pendidikan.

CONTENTS

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
CNNs	Convolutional Neural Networks
DL	Deep learning
LDA	Latent Dirichlet Allocation
LMOOCs	Language Massive Open Online Courses
ML	Machine Learning
MLP	Multilayer Perceptron
MOOC	Massive Open Online Courses
NLP	Natural Language Processing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SA	Sentiment Analysis
SVM	Support Vector Machine
TnL	Teaching and Learning

LIST OF APPENDIXES

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

CHAPTER 1

INTRODUCTION

1.1 Introduction

Sentiment analysis (SA) emerges as a specialized branch of data analytics, especially pertinent for analyzing social media content. It involves the computational identification and categorization of opinions expressed in text data to determine the writer's attitude toward a particular topic, product, or service. This approach is particularly beneficial in gauging public sentiment, tracking or monitoring brand reputation analysis, customer feedback, and understanding customer satisfaction, which is widely used in business and social media (Li, 2018; Vedavathi & KM, 2023).

Various case studies and applications underscore the versatility of SA. Similarly, SA of consumer reviews and social media posts in market research gives businesses insights into customer preferences and product reception (Peng & Jiang, 2022). For instance, in the educational sector, SA has been instrumental in evaluating online learning platforms and dissecting student feedback to enhance learning experiences and content delivery (Christian, 2022a).

These examples demonstrate the profound impact of SA in extracting meaningful information from the ever-growing expanse of social media data. This study has extensively applied sentiment and polarity analysis in online education feedback, revealing students' attitudes and preferences toward e-learning modules (Alfayoumi et al., 2021; Waheeb et al., 2022). These applications underscore the transformative potential of SA in decoding the complex tapestry of human sentiments and opinions expressed across digital platforms.

Several papers focus on SA within online education, especially during the COVID-19 pandemic. For instance, Waheeb (2022) delves into analyzing sentiments expressed by students and educators on social media platforms like Twitter. This study highlights the importance of understanding sentimental responses to online learning environments and adapting educational strategies accordingly. The other study from Li (2018) presents an innovative approach where SA is used to enhance the efficiency of language learning platforms. By analyzing the polarity of learner feedback (positive, negative, neutral), the system can adapt and personalize the learning experience, proving beneficial for both learners and educators.

In addition, SA evaluates public opinion on massive open online courses (MOOCs) (Onan, 2021; Peng & Jiang, 2022; Zhou, 2022). Also, Onan (2021) used advanced text mining and deep learning techniques to analyze reviews and feedback, providing insights into user satisfaction and areas for improvement in MOOC offerings. Peng & Jiang (2022) confirm that SA offers valuable insight into the learning experience of the student environment. SA, also used by Alfayoumi (2021), employs SA to gauge student attitudes toward online education in Jordan. This analysis helps understand the challenges and opportunities in implementing e-learning solutions in higher education. Usart (2023) explores the sentimental climate of online teacher education programs. The study evaluates the impact of online learning environments on the sentimental well-being of future teachers.

By analyzing the text's tone and context, SA helps extract subjective information, offering insights into the sentimental undercurrents of social media discourse (Deraman et al., 2021; Madani et al., 2020; Vedavathi & KM, 2023). SA employs natural language processing (NLP) methods to detect and classify sentiments expressed in written language, aiming to ascertain if these sentiments are positive, negative, or neutral, typically through ML techniques. However, its accuracy can be challenged by the nuances and complexities of human language, such as sarcasm and context.

1.2 Research Background

The challenges in SA within educational contexts are twofold: (a) data-related issues and (b) the implementation and accuracy of analysis models. Firstly, the scarcity of analytical content and historical data and the nature of the data, being noisy, imbalanced, and unstructured, pose significant hurdles regarding data quantity. Secondly, the complexities of designing and executing accurate analysis models are exacerbated by linguistic nuances and cultural variations, which are crucial elements in educational settings. These challenges impact both the data quality and the SA models' effectiveness, affecting the accuracy and reliability of SA in educational contexts (Shaik et al., 2023; Bakar, M. F. R. A., 2020).

The ability to accurately detect and interpret these sentiments remains a formidable task, necessitating the development of advanced techniques and models for practical interpretation and analysis in educational SA. In education, three unique studies unravelled the sentimental tapestry of students. Shafana & Safnas (2022) conducted the first case study, discussing sentimental classification in Jordan. They analyzed SA in online education and explored the impact of unreliable internet on student sentiments, hindered by a scarcity of deep, qualitative data. The second case was about the sentimental interpretation of students participating in Massive Open Online Courses (MOOCs) in China (Mrhar et al., 2021). The case study sought to connect students' sentiments with their learning experiences yet faced the challenge of limited data linking sentiments to course specifics. Finally, The third case was about predicting student mental stress in the context of online learning during the COVID-19 pandemic in India; a study using predictive models like support vector machine (SVM) and random forest (RF) illuminated key mental stressors among students but was constrained by the narrow scope of questionnaire responses (Nanda et al., 2022).

Each study, like a piece in a puzzle, offered glimpses into the complex sentimental landscape of online learning, which might not encompass the full spectrum of mental health challenges students face in online learning environments. The case studies collected data from Twitter and Facebook and reviewed dataset platforms (Course metadata, Coursera), where no study used text messenger tools (like Telegram or WhatsApp) for teaching and learning (TnL) activities for sentiments.

Using text messenger tools such as Telegram for TnL activities presents unique challenges in a multilingual context, particularly involving a mix of Malay and English. A significant problem arises from the prevalence of 'textism' the use of abbreviated forms of words or phrases (for example: “Dr.”= “Doctor”, “En.”= “Encik”, “Blh jum esok” = “Boleh jumpa esok”, “IDK ke mana dia pergi”= I don’t know, where he went) and the frequent incorporation of emoticons emojis, and stickers (Dewan Bahasa dan Pustaka, 2008). While enhancing expressiveness and engagement in informal communication, these elements can introduce ambiguity in an educational setting.

This situation necessitates a nuanced approach to understanding and interpreting the hybrid language of textism and multilanguage sentences in TnL activities. There is a need to develop strategies or tools that can effectively manage and make sense of this uniqueness, ensuring that the educational objectives of communication are met without losing the essence of multilingual and multimodal interactions. Thus, SA has been extensively studied across various domains, including medical, engineering, health, and education. Education influences the teaching process and student learning experiences. SA on text leverages NLP techniques to classify text into positive, negative, or neutral categories (Çeliktuğ, 2018; Medhat et al., 2014). Analyzing the tone and sentiment behind user-generated content helps to understand the sentiment, which is challenging in SA. In this context, its focus will be on SA, which is more related to research regarding educational settings.

1.3 Research Problem Statements

As emphasized in numerous studies, the field of SA within the context of education is riddled with complexities and obstacles. A primary concern is data-related issues, encompassing both the data's availability and nature. The scarcity of analytical content and historical data poses significant hurdles (Ali, 2021; Althagafi et al., 2021; Chen et al., 2019; Ji & Fangbi, 2021; Jojoa et al., 2022). Additionally, the data's nature is often characterized as noisy (Althagafi et al., 2021), imbalanced (Ji & Fangbi, 2021; Mujahid et al., 2021), and unstructured (Dina et al., 2021; Waheeb et al., 2022), further complicates analysis efforts. These issues are about data quantity and structure.

Another significant challenge lies in the implementation and accuracy of analysis models. The complexities in designing and executing these models are substantial, often exacerbated by linguistic nuances and cultural variations essential to consider in educational contexts (Imani & Montazer, 2019; Jojoa et al., 2022; Mrhar et al., 2021; Singh et al., 2021; Wang & Hu, 2022). Ensuring the accuracy of these models remains a pressing concern, given the impact of incorrect analysis on educational decisions and outcomes (Christian, 2022b; Lubis et al., 2019; Wang & Shi, 2021).

Furthermore, case studies in education reveal distinct challenges related to students and educators. The diversity and complexity of language in educational settings add another dimension to these challenges (Althagafi et al., 2021; Mrhar et al., 2021; Prinsloo et al., 2019). This is particularly evident when dealing with languages with rich morphologies or are underrepresented in research, posing significant barriers to effective SA.

Education plays a vital role in the TnL process. Instant messaging apps on smartphones have led to new ways of writing, often using unique spellings called textisms (Gómez-Camacho et al., 2023). Recent studies on various languages have analyzed WhatsApp chats between adolescents and young people to understand how the use of textisms affects their overall linguistic skills (Finkelstein & Netz, 2023; Gómez-Camacho et al., 2023; Verheijen & Spooren, 2021). As a result, several other scholars continued to study the effects of textism on literacy in the coming years (Saberia, 2016). In educational contexts, text conversation detects significant challenges for textism that hinder its effectiveness and accuracy.

Furthermore, the difficulty in accurately detecting and interpreting sentiments in educational contexts necessitates developing more advanced techniques and models for interpretation and analysis in educational SA. This concept involves understanding students' reactions, like excitement or anxiety, impacting their learning and teachers' experiences affect their teaching effectiveness. Recognizing and addressing these aspects is essential for creating effective learning environments and improving educational outcomes (Alfayoumi et al., 2021; Kaur et al., 2022; Lubis et al., 2019; Suwal & Singh, 2018).

In educational settings, there is a particular challenge in recognizing from the text through indirect parameters such as response times, which are vital for understanding student engagement and effective mental health monitoring (Alfayoumi et al., 2021; Kaur et al., 2022; Lubis et al., 2019; Suwal & Singh, 2018). These challenges include a deficiency of sufficient training data leading to inadequate representations of sentimental nuances (Chen et al., 2019), difficulties in processing and structuring unstructured data (Dina et al., 2021; Waheeb et al., 2022), and complexities in implementing analytical models that are sensitive to the subtleties of sentimental expressions, including grammatical and cultural variations (Imani & Montazer, 2019; Mrhar et al., 2021; Singh et al., 2021). Furthermore, achieving accurate recognition is complicated by the inherent complexity of different analysis scenarios and the limitations of current algorithmic classification methods (Christian, 2022b; Wang & Shi, 2021).

1.4 Research Objectives

i. To identify relevant features related to pre-processing text messenger dataset for teaching and learning (TnL) activities.

ii. To develop a model for sentiment analysis to classify students' sentiment for TnL activities based on the BERT model.

iii. To evaluate the proposed model using accuracy, precision, recall, and F1-score.

1.5 Research Questions

i. What are the relevant features for pre-processing text messenger datasets related to TnL?

ii. How to pre-processing text messenger datasets related to TnL activities?

iii. How to develop a sentiment analysis model to classify students' sentiments during TnL using text messenger dataset?

iv. What are the results of classifying student sentiment for TnL?

1.6 Research Scope

This research focuses on developing and evaluating an SA model to enhance TnL activities using data from WhatsApp and Telegram. The study addresses the challenges of noisy, unstructured, and multilingual datasets, as well as English and Malay text, incorporating features like textisms and emojis. Pre-processing techniques such as tokenization, stopword removal, stemming, and translation were applied to improve data quality. The SA model, developed using BERT, classifies student sentiments into positive, negative, and neutral categories (Onan, 2021). Its performance is evaluated using accuracy, precision, recall, and F1-score and compared with traditional ML models to demonstrate its effectiveness in educational contexts.

1.7 Research Significance

The significance of this research lies in its contribution to the advancement of SA techniques for interpreting and understanding sentiments within educational contexts, specifically through the analysis of text messenger data. This research is pivotal in several key areas. SA is effectively used for analyzing student sentiments by text in educational contexts, as demonstrated in studies by (Alfayoumi et al., 2021 Onan, 2021; Usart et al., 2023; Waheeb et al., 2022).

1.7.1 Student and Educator

The model offers substantial advancements in educational insights by accurately classifying students' sentiments during TnL activities. This more profound understanding of sentimental dynamics is essential for creating learning environments that are more effective and sentimentally supportive, thereby enhancing the overall educational experience. The research stands out for its comprehensive approach to enhancing SA in educational contexts, offering academic and practical benefits while advocating for ethical and responsible data usage. Its implications extend beyond the immediate field of education, providing relevant insights and tools across multiple disciplines.

1.7.2 SA Practitioners

The research addresses the inherent complexities of processing and analyzing text messenger data, often characterized by its unstructured, noisy, and multilingual nature. Tackling these challenges is crucial for improving the accuracy and reliability of SA in educational settings where traditional data analysis methods may fall short. In addition, the research strongly emphasizes the responsible application of data analytics tools, addressing privacy concerns and ethical dilemmas around data usage in SA. This aspect is particularly crucial in sensitive areas like education, where the ethical implications of data usage are as significant as the insights gained from the data itself.

1.8 Operational Definition

Some terms and definitions may pose clarity challenges to certain readers, thus necessitating explanations to aid comprehension. Hence, this segment aims to elucidate specific terms and definitions utilized in this research, as delineated in the table provided.

Table 1.1

Operational Definition

List of terms and abbreviations	Definitions
Machine Learning (ML)	Machine learning is like teaching computers to learn from examples and patterns instead of giving specific instructions for every task.

List of terms and abbreviations	**Definitions**
Sentiment Analysis (SA)	SA is a process where computers analyze text to determine its sentimental tone. It helps identify whether the text expresses positive, negative, or neutral sentiments.
Massive Open Online Courses (MOOCs)	MOOCs are online courses designed to be open to anyone with internet access and typically offer unlimited participation and access to course materials.
Natural Language Processing (NLP)	NLP is a field of artificial intelligence (AI) that focuses on enabling computers to understand, interpret, and generate human language meaningfully and usefully.
Support Vector Machine (SVM)	SVM is a supervised machine learning algorithm used for classification and regression tasks.
Language Massive Open Online Courses (LMOOCs)	LMOOCs are online courses that cater specifically to language learning. They typically offer various language learning materials, including lessons, exercises, multimedia content, and interactive activities.

1.9 Summary

This chapter presents a comprehensive overview of the importance and challenges of SA in educational contexts, emphasizing its role in interpreting and understanding sentiments through text messenger data. It highlights the significance of data in various fields, focusing on social media as a rich data source. The research aims to innovate SA techniques to address these challenges, proposing a model to classify students' sentiments during TnL activities accurately. Additionally, it emphasizes the ethical aspects of data usage. The significance of this research lies in its potential to enhance educational insights, improve the reliability of SA, and contribute practically and theoretically to the field. Literature review discussions will be covered in the next chapter, Chapter 2.