EMOTION RECOGNITION USING LARGE LANGUAGE MODEL

A Thesis Submitted
In Partial Fulfillment of the Requirements
for the Degree of

MASTER OF TECHNOLOGY

In

Artificial Intelligence

By

VERSHIKA SRIVASTAVA
(Roll No. 2K22/AFI/28)

Under the Supervision of
DR. SANJAY KUMAR
(Asst. Prof, Department of Computer Science & Engineering)

To The
Department of Computer Science and Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042. India

May, 2024

ACKNOWLEDGEMENTS

I am highly indebted to **Dr. Sanjay Kumar** for his guidance and constant supervision as well as for providing necessary information regarding the project & alsofor his support in completing this research work. I would like to express my gratitudeto the **Head of the Department (Computer Science and Engineering, Delhi Technological University)** for their kind cooperation and encouragement whichhelped me in the completion of this research work. I would like to express my specialgratitude and thanks to all the Computer Science and Engineering staff for giving me such attention and time. My thanks and appreciation also go to my colleagues and the people who have willingly helped me with their abilities.

Vershika Srivastava

2K22/AFI/28

Department of CSE

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I, Vershika Srivastava, Roll No. 2K22/AFI/28 student of M.Tech (Artificial Intelligence), hereby certify that the work which is being presented in the thesis entitled "**Emotion Recognition using Large Language Model**" in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Artificial Intelligence in the Department of ComputerScience and Engineering, Delhi Technological University is an authentic record of mywork carried out during the period from August 2022 to June 2024 under the supervision of Dr. Sanjay Kumar, Asst. Prof, Department of Computer Science and Engineering.The content presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Place: Delhi **Candidate's Signature**

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE

Certified that **Vershika Srivastava** (Roll No. 2K22/AFI/28) has carried out the research work presented in the thesis titled "**Emotion Recognition using Large Language Model**", for the award of Degree of Master of Technology from the Department of Computer Science and Engineering, Delhi Technological University, Delhi under my supervision. The thesis embodies the result oforiginal work and studies carried out by the student herself and the contents of thethesis do not form the basis for the award of any other degree for the candidate or submission from any other University /Institution.

Dr. Sanjay Kumar
(Supervisor) Department of CSE
Date: Delhi Technological University

Emotion Recognition Using Large Language Model

ABSTRACT

Emotion recognition is the process of recognizing human emotions from facial expressions, voice patterns, or text, utilizing various techniques like Machine Learning or Natural Language Processing. It is a crucial aspect of human interaction that has drawn a significant amount of attention, particularly with the emergence of natural language processing (NLP) techniques. This paper explores the use of transformer based large language models in the field of emotion recognition due to their ability to capture complex contextual relationships in data. Large Language Models are an important development in the field of emotion recognition as they can accurately analyze and interpret complex human emotions from textual data. This innovation has enhanced the context-sensitivity of emotional analysis, opening new possibilities in various fields like mental health care, and customer services, where understanding complex emotional states is crucial. This research provides an overview of the current development and approaches of emotion recognition using Large Language Models (LLMs). It explores the recent methodologies in LLM-based emotion recognition, emphasizing the models' unparalleled capabilities to understand and interpret the complex emotions in text. It explores how well transformer architectures adeptly capture and model the nuances of emotion in textual data. Through extensive experimentation and comparative analysis, this research evaluates various model's performance in precisely identifying and classifying emotions. The research further delves into the challenges and limitations faced by current LLMs in emotion recognition. Additionally, addresses the research gaps and outlines possible future paths. The outcome of this work improves the understanding of LLM based techniques for emotion recognition and provides important new information about their usefulness in various real-world contexts.

LIST OF RESEARCH PAPERS

1. Vershika Srivastava and Sanjay Kumar, "**Emotion Recognition Using Large Language Models: An Overview**" accepted and presented in '3rd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (IEEE ICPEICES-2024)' at DTU in April 2024.

2. Vershika Srivastava and Sanjay Kumar, "**Transformer Based Techniques for Emotion Recognition**" accepted and presented at '4th International Conference on Machine Learning and Big Data Analytics (ICMLBDA-2024)' at NIT Kurukshetra in May 2024.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Overview

Understanding the emotional undercurrents of written communication is paramount for achieving natural and engaging human-computer interaction. This ability holds immense potential for transforming fields like customer service, mental health support, and targeted content delivery. Traditionally, emotion recognition relied on methods that were limited in scope, often utilizing sentiment dictionaries or rudimentary machine learning algorithms. However, a new era has dawned with the rise of deep learning, particularly the emergence of Large Language Models (LLMs) [1]. LLMs, trained on vast troves of text and code, possess an exceptional ability to decipher complex human emotions within textual data. These powerful models, equipped with sophisticated architectures and fueled by massive datasets, achieve superior results compared to traditional approaches. This thesis delves into the potential of LLMs to revolutionize the field of emotion recognition from text [2]. We will embark on a journey that explores the limitations of earlier methods, unveils the capabilities of LLMs, and analyzes their transformative impact across various domains. By venturing into this exciting frontier of research, we aim to contribute to the development of more refined and nuanced methods for computers to understand the emotional tapestry woven within human text [3].This thesis delves into the world of LLM-based emotion recognition, exploring its potential and limitations [4]. We will examine how these advanced models, capable of complex interpretations akin to human comprehension, are transforming the way we analyze emotions in text. By drawing insights from leading research in this field, we aim to provide a comprehensive understanding of this rapidly evolving technology.

The Large Language Models have wide range of application including sentiment analysis. LLMs have reflected promising performance domain specific emotion recognition like finance and health [5]. The ability to find valuable insights from the texts for emotion recognition by LLMs has improved the decision-making processes. The potential for recognizing emotions in a diverse range of situations makes the models versatile across various contexts. In recent years, various deep learning based techniques has been pivotal in solving many real-life and business problems such emotion recognition, medical imaging, social network analysis, computer vision, and many others [5-10].

Furthermore, LLMs possesses fascinating capability in mimicking human-like emotional interactions. This can be helpful in various applications where understanding and responding to emotional cues is crucial, such as customer service [11].

This research will comprehensively explore the application of LLMs for emotion recognition. We will delve into their cutting-edge capabilities, analyze the challenges and ethical considerations they present, and finally, offer an insightful overview of the current state of this field. By examining both the potential and limitations of LLM-based emotion recognition, we aim to contribute to its future development and responsible application across various domains.

1.2 Motivation

The amazing progress in AI and language processing has fueled research into using large language models (LLMs) to understand emotions in text. These models are performing better at figuring out complex feelings from text data. This has huge potential in areas like finance, customer service, and anywhere else where understanding emotions can improve interactions and decision-making. However, with any new technology, there are challenges and questions to consider. This report will explore both the exciting possibilities and the issues we need to address as LLM emotion recognition keeps developing. Through this study, we aim to explore:

1) Various large language models and approaches for emotion recognition
2) Datasets availability for training of the models
3) Applications of emotion recognition utilizing LLMs in various domains
4) Challenges occurring while performing emotion recognition with LLMs

1.3 Objectives

The primary objective of conducting emotion prediction using Large Language Models is to understand human emotions better from written text using powerful language models. This research aims to explore and develop robust and accurate techniques for emotion detection, classification, and prediction, and eliminating the limitations of traditional sentiment analysis. This could be a game-changer for fields like mental health support, understanding customer feelings, and tracking trends on social media. Additionally, this study seeks to explore the potential of Large Language Models to provide personalized recommendations and interventions based on an individual's emotions, ultimately enhancing well-being and user experiences in applications across domains. Through an investigation of methodologies, data sources, and model architectures, this research endeavors to unlock the transformative power of Large Language Models in deciphering the complex tapestry of human emotions, offering novel insights and practical applications.

1) To conduct an extensive literature review on the recent research in the field of emotion recognition, specifically focusing on the application of large language models.
2) To implement a LLM-based model for human recognition and classification using textual data.
3) To validate how effectively the model works using different performance evaluation metrics like accuracy, precision, recall, and F1-Score and how it performs stands up against the traditional methodologies.

1.4 Thesis Organization

The thesis is structured into six distinct chapters as follows:

- Chapter 1 covers the introduction of the topic, outlines the motivation behind the research work, and sets forth the objectives of the study.
- Chapter 2 offers a comprehensive review of the existing literature on methods for emotion recognition using Large Language Models. It also highlights the limitations of these methods and examines the datasets

employed in prior research works.

- Chapter 3 is dedicated to the development of the model, describing the implementation process.
- Chapter 4 evaluates the results achieved and discusses the hyper-parameters used in the experimental framework.
- Chapter 5 concludes the study, discussing potential future research and the industrial relevance of this specialized field.

CHAPTER 2

LITERATURE REVIEW

2.1 Background

2.1.1 Emotion Recognition

Emotions are psychological states comprised of a variety of feelings, ideas, and physical responses to external events or stimuli. They have a significant impact on behavior, judgment calls, and interpersonal relationships throughout the human experience. Understanding emotions is essential to human existence as it is an integral part of social interaction and communication [12]. Understanding emotions facilitates the development of empathy and deep connections by enabling us to recognize and react suitably to the emotions of others. Consequently, Emotion Recognition is an important field within Natural Language Processing (NLP), which is concerned with the extraction and analysis of subjective data from various sources. Its main goal is to comprehend the emotions and views expressed by a writer or speaker regarding a specific topic or to discern the overall sentiment within a document.

2.1.2 Emotion Recognition Techniques

From lexicon-based approaches in the early phases of emotion recognition, the area of computational linguistics and artificial intelligence has advanced to the further developed and complicated use of Large Language Models (LLMs), as shown in Fig 2.1, as AI developed over time, more sophisticated methods became available. A recent major accomplishment in this field is represented by Large Language Models. Due to their extensive training on large datasets, LLMs—like GPT (Generative Pre-trained Transformer) models—are able to process and produce language in a contextually rich and complex manner.

a) **Lexicon-based method:** Initially, lexicon-based methods were the foundation of emotion recognition. These early methods made use of lexicons with emotional valences annotated. The procedure was simple: look for these words in the text, then use their cumulative scores to determine the sentiment. This method was limited in its ability to comprehend colloquial language, sarcasm, and context.

b) **Rule-based Systems**: Rule-based systems rely on predefined rules or patterns to determine sentiment. These rules can include linguistic rules, syntactic patterns, or regular expressions. While they may lack the flexibility of machine learning approaches, they can be useful for specific domains or languages where labeled data is scarce.

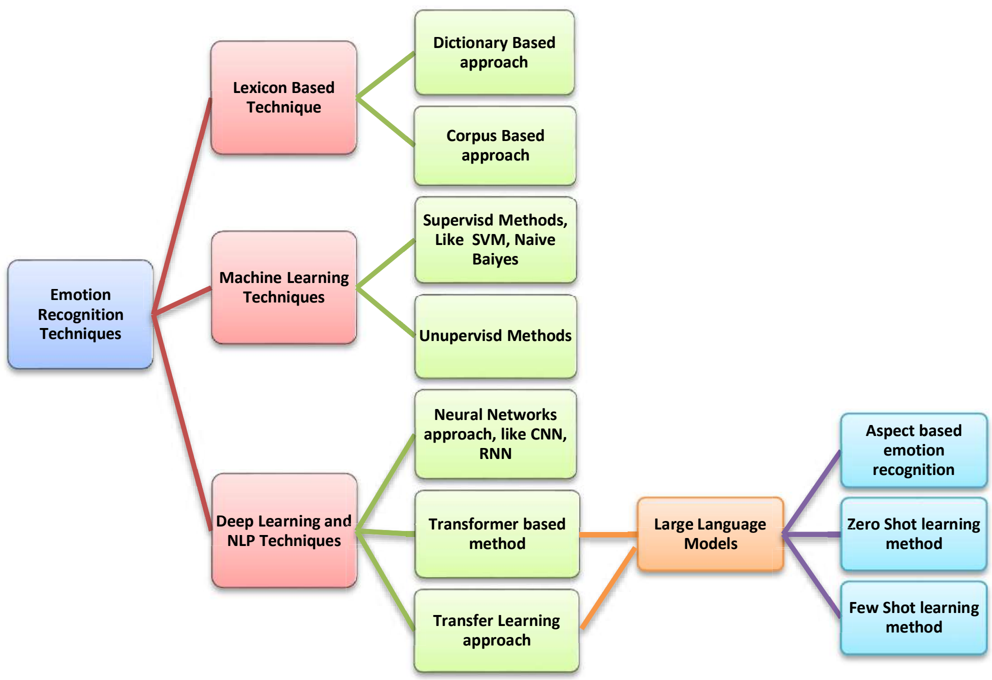


Fig.2.1 Evolution of Emotion Recognition Techniques

c) **Machine Learning:** As technology evolved machine learning—more specifically, supervised learning models became increasingly significant. Using labeled datasets, these machine learning models were trained to identify patterns corresponding to various emotions. This was a key development which made it possible to analyze text in a more complex way, but feature engineering remained crucial.

d) **Deep Learning and Natural Language Processing:** Substantial progress was made with the introduction of Deep Learning, particularly with neural networks. More sophisticated text processing could be achieved by models such as Recurrent Neural Networks (RNNs), which are capable of interpreting sequences and contextual data. This change was crucial for expressing the nuances and complexity of language [6]. With the emergence of transformers, the transformer-based models have emerged as powerful tools in sentiment analysis. These models, such as BERT (Bidirectional Encoder Representations from Transformers) and its variants, leverage self-attention mechanisms to capture long-range dependencies in text efficiently, allowing them to better understand the context and semantics of language. By fine-tuning transformer-based models for emotion recognition tasks, researchers have achieved state-of-the-art performance in accurately identifying and classifying emotions expressed in text.

- **Transformers**

Transformers are powerful deep learning models that use a special technique called self-attention [17]. This lets them focus on the most important parts of incoming data, like words in a sentence or features in an image. They're commonly used for tasks in natural language processing (like understanding text) and computer vision (like analyzing images or audio). Transformers have two main parts:

- *Encoder:* This processes the input data, uncovering its deeper meaning and relationships.
- *Decoder:* This uses the encoder's insights to generate an output, such as a translation or a description.

The key to a Transformer's success is its self-attention mechanism [18]. This allows it to pay closer attention to specific parts of the input data that are most relevant to the task at hand. By calculating each input element's relevance in relation to each output element, this method scores each pair of elements. Higher scores show that the model is paying more attention. Several types of the attention mechanism can be used: self-attention, in which the model focuses on its own input or output; and cross-attention, in which the model simultaneously evaluates its input and output. Fig. 2.2 shows the architecture of the transformer including self-attention mechanism.

Early advancements in AI combined deep learning with Natural Language Processing (NLP). Combining these techniques led to better understanding of language structure (parsing), meaning (semantics), and even the ability to generate human-like text. This progress was crucial for unlocking the subtleties of human language, particularly emotions.

Today, generative AI and large language models (LLMs) are the cutting edge. LLMs, like GPT (Generative Pre-trained Transformer) models, are a major leap forward in language comprehension [7-8]. These models are trained on massive amounts of text data, allowing them to create content that is both coherent and relevant to the situation. LLMs excel in various NLP tasks, including recognizing emotions. They can effectively capture the complexities and nuances of human communication.

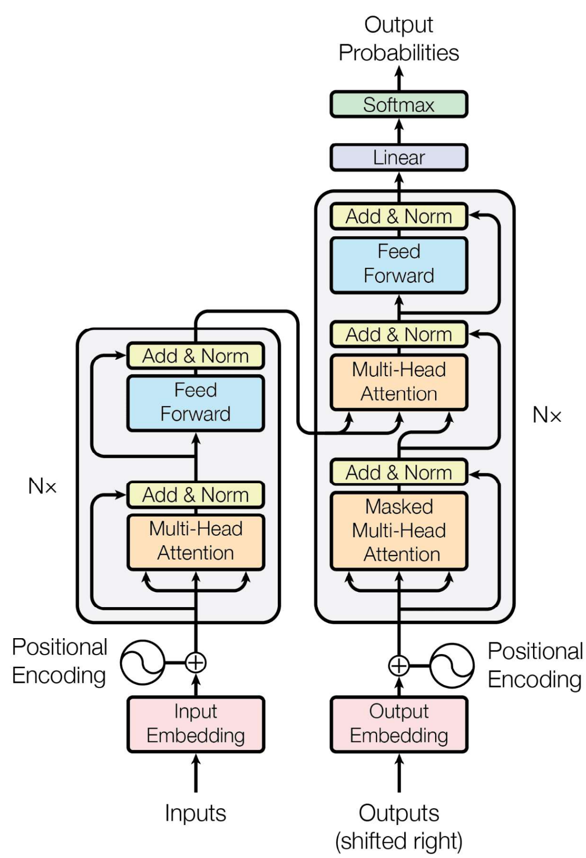


Fig.2.2 Architecture of Transformer [18]

- **Large language models**

Large language models (LLMs) are game-changers for natural language processing (NLP). They excel at understanding complex language patterns, learning effectively, and drawing from massive stores of information. These models are remarkably good at understanding context and representing difficult ideas. They use statistical methods and computer algorithms to analyze and generate text that's almost indistinguishable from human language.

The secret sauce behind LLMs is their size and sophistication. The more parameters an LLM has (like GPT-3's mind-boggling 175 billion!), the better it can mimic human brain functions when creating text. OpenAI, a leader in AI research, is a key player in developing these powerful models. Their GPT-3 model, for example, uses a special architecture called a transformer to produce text that's incredibly close to human language. Since GPT-3 was trained on a vast amount of internet data, it can generate coherent and relevant phrases, even if its performance can vary depending on the task.

The potential applications of LLMs are vast and ever-growing. Their ability to generate human-quality text makes them useful for a variety of tasks, including machine translation, summarizing text, creating chatbots, and even understanding complex emotions in language.

Large Language Models (LLMs) which are crucial to NLP research, are constantly evolving to handle a broad spectrum of tasks. Organizations leading the way in this breakthrough are such like OpenAI, with GPT-3.5. Our methodology based on GPT-3.5 represents a major advancement, particularly in the analysis of social media information.

Overall, the area has progressed from straightforward keyword-based techniques to complex models that can comprehend and imitate human language, providing insights into the emotional states expressed through text. The progression as shown in Fig.2.1 reflects broader trends in AI and NLP, showcasing the enormous improvement in our capacity to interpret and comprehend human language and emotions.

2.2 Literature Survey

Emotion recognition using Large Language Models (LLMs) is an emerging field that has garnered significant attention and research across a range of approaches, uses, and theoretical frameworks, which has resulted in significant advancements in artificial intelligence and natural language processing. This survey reviews the application of large language models (LLMs) and cutting-edge artificial intelligence (AI) to the recognition of emotions. It highlights the progress in this area but also points out the challenges and ethical issues that come with it. Based on significant research findings in this area, we try to provide an overview of the current state of emotion recognition with LLMs, including its applications, issues, and potential future developments. The research papers were searched using the subsequent criteria:

- **Time Period:** This paper explores the latest developments and uses of emotion recognition, particularly in relation to Large Language Models made in this area from 2021 to 2024.
- **Keywords:** This research employs a focused keyword search approach that includes particular terms like 'Emotion recognition', 'Generative AI', 'GPT', 'Large Language Model', 'LLM', 'NLP', 'Sentiment Analysis' and others. This approach ensures precise and thorough analysis of the most current advancements.
- **Inclusion/Exclusion Criteria:** The research exclusively included conference and journal papers written in English. The studies that explored the intersection of LLMs and emotion recognition emphasizing impactful and relevant developments in the field were prioritized.

In this survey, we look at a variety of techniques used by researchers using diverse datasets to optimize the performance of LLMs across several domains like, the study by Guven et al. [19] investigated sentiment analysis in Turkish by utilizing a number of language models, including DistilBERT, ELECTRA, BERT, and ALBERT. It presented a text filtering technique that eliminates words that run counter to the text's general sentiment in order to improve sentiment analysis. The study showed that the ELECTRA model, in conjunction with the suggested filtering procedure, obtains the best result, indicating the effectiveness of this methodology for emotion recognition. The paper presented by Üveges et al. [20] fine-tuned the BERT model (huBERT) for emotion analysis of political text in Hungarian literature. HunEmBERT, a refined model, was designed to classify sentiment as well as emotion in political communication, tackling the difficulties of identifying different emotions in a language with a complex morphology like

Hungarian. The work proved HunEmBERT's effectiveness and shed light on the particular classification flaws and restrictions of BERT-based models for language-specific sentiment and emotion recognition. Using transformer based BERT model given by Patel [5], delved into emotion recognitions in the airline sector. This study demonstrated how effectively BERT not only works but outperforms other ML based algorithm in the airline industry for processing and assessing client emotion. LLM's applicability was explored in finance sector as well, where an enhanced retrieval-based LLM framework for financial sentiment analysis was presented in a study by Zhang et al. [21]. It enhanced emotion classification accuracy by utilizing retrieval-augmentation and instruction-tuned LLMs, which outperformed traditional models and LLMs like ChatGPT and LLaMA. This method addressed the issues with directly applying LLMs to financial sentiment research by supplying more context from other sources and matching LLM predictions to user intentions. Another application of Large Language Models such as GPT-3 and PaLM for financial sentiment analysis was explored, where it used LLMs in a semi-supervised learning approach to produce imprecise financial sentiment labels for articles on Reddit [4]. The study depicted improved accuracy when LLMs were prompted with Chain-of-Thought summaries. Furthermore, the studies have also addressed the challenges with LLM for emotion recognition proposing techniques for performance optimization. Mao et al. [16] looked into the biases inherent in pre-trained language models (PLMs) when used for emotion recognition. It explored the effects of the factor like model architecture, prompt templates, and label-word selection on performance of the task. It revealed major differences in performance between several PLMs and highlighted the importance of fine-grained emotion taxonomies can have on the results of emotion recognition. As an application of LLM in e-commerce, the effectiveness of BERT and RoBERTa in conjunction with the GPT-3.5 and LLaMA-2 models was compared for emotion recognition in e-commerce product reviews. The study done by Roumeliotis et al. [23] attempted to investigate the influence of LLMs on customer satisfaction analysis as well as the ability LLMs in comprehending consumer satisfaction in e-commerce.

Table 2.1 Literature Survey for Emotion Recognition using LLM

Authors, Year	Description	Model Used	Dataset	Performance	Pros	Cons
Rui Mao et al. , 2023 [13]	Studied the biases of pre-trained language models in prompt-based sentiment analysis	BERT, RoBERTa, ALBERT, BART	Amazon product reviews for electronic devices.	RoBERTa model was more effective in emotion detection than other BERT based PLMs.	Provides insights into biases of PLMs in affective computing tasks.	Biases in PLMs that can affect the reliability of their outputs in sentiment analysis
Xiang Deng, 2023 [12]	Presented a semi-supervised learning approach using LLM for market sentiment	GPT-3/PaLM	Reddit posts with financial content	Produced stable and accurate labels, with the final model outperforming on par with the existing supervised models	Effective in generating stable and accurate financial sentiment labels.	Ethical considerations limit its application in high-stakes areas like investment decisions
Boyu Zhang et al. , 2023 [4]	Introduced retrieval augmented LLM framework for financial sentiment analysis, integrates external knowledge retrieval and instruction tuning	LaMA 7B with a Retrieval-augmentation module	Twitter Financial News dataset and FiQA dataset	Achieved 15% to 48% performance gain in accuracy and F1 score compared to traditional models	Improved model understanding of sentiment in financial news using instruction tuning	Reliance on textual similarity for information retrieval, overlooking macro and micro economic information
István Üveges et al. , 2023 [15]	Created a fine-tuned transformer-based model for emotion analysis of Hungarian political texts.	Hungarian BERT (huBERT)	Hungarian Parliamentary speech transcripts	Sentiment classification model achieved 0.866 macro average and 0.9149 weighted average F-Score.	Contribution to the field in political communication using Hungarian literature	Challenges in detecting linguistic phenomena like irony or sarcasm in Hungarian language.

Zekeriya Anil Guven et al. , 2022 [14]	Compared various pre-trained language models for emotion recognition in Turkish language, introduced text filtering method to enhance sentiment analysis accuracy.	BERT, DistilBERT, ELECTRA, and ALBERT specifically trained for the Turkish language.	Turkish hotel and movie review datasets from Hacettepe University.	BERT model achieved highest accuracy of 98.38% for hotel dataset, while ELECTRA model 92.21% for the movie dataset	Robust against reduction in training data	The method's effectiveness is depend on the specific characteristics of the dataset, like prevalence of opposing sentiment words
Konstantinos I. Roumeliotis et al., 2024 [16]	Evaluated effectiveness of various models for sentiment analysis in e-commerce.	GPT-3.5, LLaMA-2, BERT, RoBERTa	Product reviews dataset from Amazon and EBay	After fine-tuning process, the GPT-3.5 model exhibited better performance to the LLaMA-2 model by a margin of 2.39% in predicting star ratings for product reviews.	High accuracy post fine-tuning	Conclusions are based on a randomly selected sample of product reviews, which do not fully represent the diverse e-commerce landscape.
Akash Patel et al., 2023 [5]	Studied emotion recognition on customer feedback for an airline using machine learning and transformer based techniques.	BERT, Naive Bayes, SVM, Decision Tree, Random Forest, Adaboost	Airline reviews dataset from Kaggle	BERT model outperformed traditional ML models eith an accuracy of 83%	Application of emotion recognition in Airline Sector with high performance using transformer based models like BERT	Limitations in text generalization and translation

2.3 Dataset Studied

In the survey, we explored the diverse range of datasets that have been used in the studies to implement LLMs for emotion recognition, these datasets play important role in developing the field of emotion recognition across multiple domains. This section provides a summary of the datasets that are currently available and have been majorly utilized by the researchers for emotion recognition task using LLMs. Table 2.2 depicts the dataset accompanied by the features and emotion categories that were employed in the studies displayed in Table 2.1.

Table 2.2 Datasets covered in this Literature Survey

Dataset	Emotions category	Total count
The data were sourced from Twitter, proposed on WASSA-2017 Shared Task on Emotion Intensity [17].	Anger, Fear, Joy, Sadness, Ambiguous	7097
Reddit posts with financial content	Positive, Neutral, and Negative	20000
Twitter Financial News dataset	Bearish, Bullish, or Neutral. 11928	9540
FiQA	Positive, Neutral, and Negative	961
Hungarian Parliamentary speech transcripts (2014-2015)	Neutral, Fear, Sadness, Anger, Disgust, Success, Joy, Trust	19002
Turkish hotel review dataset proposed Hacettepe University	Positive, Neutral, and Negative	53400
Turkish movie review dataset proposed Hacettepe University	Positive, Neutral, and Negative	11600
Product reviews dataset from Amazon and EBay of 616 distinct products	Positive, Neutral, and Negative	5029
Airline reviews dataset from Kaggle	Positive, Negative Neutral	14768

2.4 Limitations in Existing Works

a) Social Bias: Large amounts of text data are used to train the models, which may unintentionally introduce the existing biases from the training set into the model's results [24]. Such biases can perpetuate and amplify societal prejudices, leading to unfair or skewed results. Addressing social bias in models is essential for promoting fairness, inclusivity, and unbiased decision-making.

b) **Working with Multimodal Data:** Subsequent research ought to investigate the incorporation of text-based emotion recognition with different data formats, like auditory and visual inputs. This multimodal method, which integrates verbal cues with non-verbal ones like tone of voice or facial expressions, can result in a more detailed understanding of emotions by offering a richer and more comprehensive picture of emotional states [25].

c) **Working with Multilingual Data:** As per the observation, most of the application of LLMs used majorly English language, which narrows the opportunity to implement LLMs at various tasks [26]. Deficiency of multilingual text corpus is also one of the factors for the same.

CHAPTER 3

PROPOSED METHODOLOGY

In this chapter we discuss about the proposed methodology for the emotion recognition for textual data using Large Language Models. Here, we have utilized pre-existing labeled datasets gathered from various social media platforms available on kaggle and used state-of-the-art BERT, ALBERT and ROBERTA models to for the emotion recognition task. In the upcoming sub-sections we have discussed about overall process of proposed including data acquisition, pre-processing, model training and validation. The Fig 3.1 shows the overall process of proposed methodology.

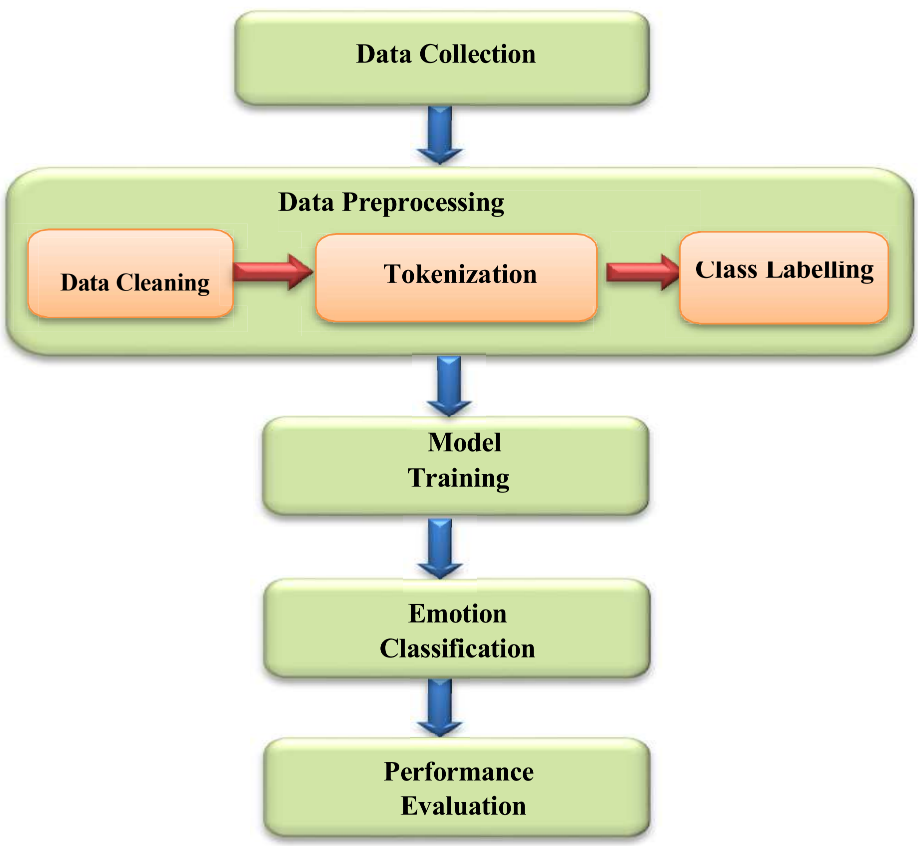


Fig 3.1 General workflow of proposed methodology

3.1 Data Acquisition

In the data collection step for emotion recognition task, we have used pre-existing labeled datasets gathered from various social media platforms [30] available on kaggle which consist of nearly 58000 entries. These datasets contain samples of text, paired with corresponding emotion labels such as positive, neutral and negative etc.

3.2 Data Validation and Cleaning

The data validation process involves checking the standards of the data for the research ensuring its completeness. The cleaning step involves checking whether the collected data error free and do not consist of incorrect or irrelevant entries. This step ensures that only high quality data is utilized for the model training for achieving better performance. Post validation and collection, the final dataset undergoes through different data preprocessing steps which are discussed in next sub-sections.

3.3 Data Preprocessing

Data preprocessing is an important phase in preparing the raw data for training the model for emotion recognition task. The data preprocessing stages for emotion recognition is broken down as follows:

3.3.1 Text Cleaning:

Raw text data contained noise in the form of special characters, punctuation, emoji, and non-standard encoding. To ensure that the text is in a clear and consistent structure, text cleaning was done involving elimination of the extraneous components. This included removing hashtags, mentions and links, and converting text to lowercase, and eliminating punctuation.

3.3.2 Text Analysis:

After cleaning the raw data, the text data was analyzed using various techniques, such as creating visual graphs and plots to assess various aspects of the text. This helped to clean all the irrelevant data. Such as, finding out texts where the word count was below 10 and removing them considering their relevance. Fig. 3.2 shows the distribution of texts having word count less than 10.

3.3.3 Tokenization:

Tokenization is the process of dividing the text into smaller pieces known as tokens. Tokens in the context of natural language processing (NLP) are usually characters or words. Since transformer models function at the token level the input text was tokenized.

3.3.4 Padding and Truncation:

BERT requires fixed-length input sequences. If the input text is longer than the maximum sequence length supported by BERT, it needs to be truncated. Conversely, if the input text is shorter, it needs to be padded with special tokens to match the maximum sequence length.

3.3.5 Label Encoding:

Sentiment labels (e.g., positive, negative, neutral) need to be encoded into numerical format suitable for model training. This may involve assigning numerical values to each sentiment class (e.g., 0 for negative, 1 for neutral, 2 for positive).

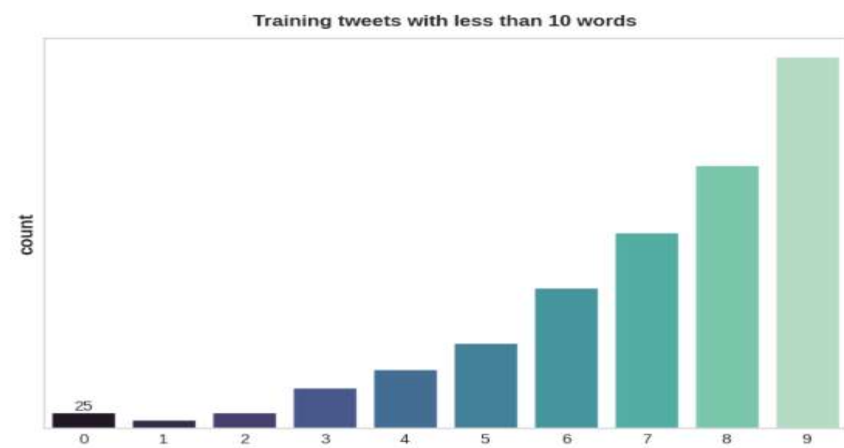


Fig. 3.2 (a) Count of tweets having word count less than 10 in Training set

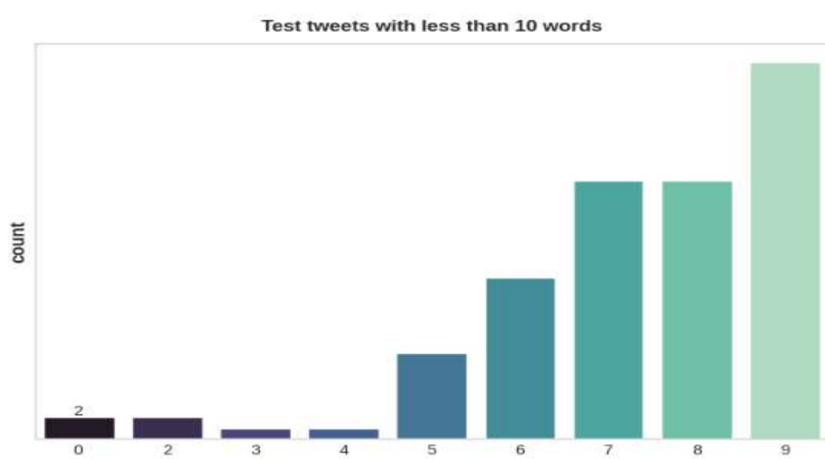


Fig. 3.2 (b) Count of tweets having word count less than 10 in Test set

3.3.6 Class Balancing:

In order to achieve better performance of the models without being impacted by any bias, we checked whether the dataset is evenly distributed into the three categories of the emotions. As the dataset was found to be unbalanced, we used Random Oversampling technique to resulting in equal data count for each class: ‘Positive’, ‘Neutral’, and ‘Negative’. Fig. 3.3 shows the distribution of data before and after class balancing.

```
df['Sentiment'].value_counts()

2    17999
0    15364
1     7560
Name: Sentiment, dtype: int64
```

.

Fig. 3.3 (a) Data Distribution before balancing

```
train_os['Sentiment'].value_counts()

2    17999
0    17999
1    17999
Name: Sentiment, dtype: int64
```

Fig. 3.3 (b) Data Distribution after balancing

3.3.7 Train-Test Split:

Distinct training and testing sets are created from the preprocessed data. The testing set is used to assess the model's performance on untested data, whereas the training set is used to train the model. In addition, a validation set is made to adjust the hyper-parameters while training the model. The distribution of training, validation and testing dataset is given in table for our dataset.

Table 3.1 Final Distribution of Training Dataset, Validation dataset, and testing dataset

Dataset	Count
Training Data	48597
Validation Data	5400
Testing Data	3793

3.4 Model Architecture Selection

In this study, we have selected transformer based BERT model and its variant to implement emotion recognition task. Depending upon the computational resources available, we have used BERT, ROBERTA and ALBERT model, also Naïve Bayes machine learning algorithm is used as base to compare the performance of the models with traditional techniques.

3.4.1 Transformer Models

Transformer models are a new type of architecture for understanding language. They're different from older models because they can look at entire sentences at once, instead of needing to process words one by one [28]. This makes them faster to train and allows them to better understand how words relate to each other, even if they're far apart in the sentence. Here's what makes transformers special:

i. **Self-attention:** This is a superpower that lets transformers focus on the most important words in a sentence, depending on what other words are around them. It's like being able to listen to a conversation and pay attention to the speaker who matters most at any given moment.

ii. **Process whole sentences at once:** Unlike older models that take turns looking at each word, transformers can consider all the words together. This is like being able to read an entire sentence instead of just one letter at a time.

Because of these features, transformers are really good at understanding complex language and are used in many powerful NLP models like BERT, GPT, and T5. These models can do things like translate languages, understand emotions in text, and even write new text themselves.

a) Bidirectional Encoder Representation from Transformers

BERT is a powerful language model created by Google that's really good at understanding the meaning of words in text [27, 30 34, 37]. It transformed the NLP tasks by applying attention approaches and transformer architecture. It's like a super-smart reader that can not only understand individual words but also how those words connect to each other in a sentence. Here's what makes

BERT special:

i. **Attention superpower:** BERT uses something called "self-attention" to focus on the most important words in a sentence, depending on the other words around them. Imagine you're reading a paragraph - BERT can pay close attention to the key words and how they relate to each other, just like you would to understand the main points.

ii. **Bidirectional reading:** This way, it can understand how words at the beginning of a sentence connect to words at the end, which is crucial for getting the full meaning. BERT consists of multiple layers of transformer encoders. After conducting self-attention on the input embeddings, or word representations, each encoder layer feeds the attended representations through feed-forward neural networks. BERT has the ability to capture multiple levels of contextual information, ranging from syntactic to semantic, because of the self-attention mechanism in each layer [32].

iii. **Pre-training on tons of text:** BERT is trained on extensive text corpora in the pre-training phase through two unsupervised learning tasks: next sentence prediction (NSP) and masked language modeling (MLM). In MLM, a model is trained to predict masked words in the input text by using the context provided by the surrounding words. By using pre-training, BERT acquires rich contextual representations that can be fine-tuned for tasks like text classification, emotion recognition, and question answering with less task-specific training data [40].

Because of these features, BERT can be fine-tuned to perform many kinds of NLP tasks, like classifying the sentiment of text, recognizing emotions in writing, and even answering your questions

b) Robustly Optimized BERT Pre-training Approach

RoBERTa (Robustly optimized BERT approach) is a fine-tuned version of BERT, created by Facebook AI [35]. It takes the core ideas of BERT's architecture and pre-training, but adds some improvements to make it more robust (better performing) and faster. There are a few key technical differences between BERT and RoBERTa, which are mentioned below:

i. **Masking strategy:** BERT uses a static masking approach, where the same tokens are masked throughout the entire pre-training process. RoBERTa uses dynamic masking. In each training epoch, different tokens are masked on the fly. This exposes the model to a wider variety of masked tokens and helps it learn better from the massive amount of unlabeled text data. And, instead of employing the next-sentence prediction job that BERT included, RoBERTa refines the hyper-parameters and eliminates them, concentrating only on masked language modeling (MLM) and improving the training objectives [11].

ii. **Batch size:** RoBERTa uses larger batch sizes during pre-training. Batch size refers to the number of training examples processed together. This allows RoBERTa to take advantage of parallelism better, meaning it can utilize more computing power to train faster. Imagine BERT training on one exercise bike at a time, while RoBERTa has access to a whole spin class.

Overall, these technical tweaks in RoBERTa's pre-training process help it achieve better performance and efficiency compared to BERT

c) A Lite BERT

ALBERT also known as A Lite BERT, is a leaner and faster version of BERT, designed for efficiency and handling bigger tasks [36]. While BERT is powerful, it can be computationally expensive due to its large number of parameters (adjustable values the model learns). There are few differences between ALBERT and BERT such as parameterization method. While BERT requires a large number of parameters, ALBERT uses parameter reduction techniques such parameter sharing and factorization, which significantly lower the number of parameters without sacrificing performance. ALBERT is plays a significant role for large-scale applications since it uses less memory and can be trained faster in comparison to BERT, and since it needs fewer parameters ALBERT is more parameter-efficient [37]. The differences between both models are given below.

i. Parameter sharing and factorization: Instead of having separate embedding matrices for word embeddings and transformer encoder layers, ALBERT utilizes parameter sharing. This means it uses the same embedding vectors for both tasks. This reduces redundancy and lowers the number of parameters significantly [38].

ii. **Sentence-order prediction:** ALBERT adds a new training step where it predicts if two sentences follow each other logically. This helps it grasp context even better. Despite these changes, ALBERT keeps the core BERT structure:

iii. **Transformer encoder layers:** These layers process information step-by-step, allowing the model to build understanding.

iv. **Self-attention:** This lets ALBERT focus on important words based on their connection to others in the sentence, similar to how you focus on key parts of a conversation.

Overall, ALBERT achieves similar or better results than BERT on various NLP tasks, but with less processing power. This makes it a good choice for large-scale applications where efficiency is crucial. Table 3.2 compares the various aspects of the transformer-based models – BERT, ALBERT, and RoBERTa [39].

Table 3.2 Comparison of BERT, ALBERT, and RoBERTa

Parameter	BERT	ALBERT	RoBERTa
Architecture	Transformer	Transformer	Transformer
Pre-training	Whole Word Masking (WWM)	Sentence-order prediction (SOP)	Dynamically Masked Language Modeling (DMLM)
Training Objective	Masked Language Model (MLM)	Masked Language Model (MLM)	Masked Language Model (MLM)
Parameter Sharing	N/A	Cross-layer parameter sharing	Cross-layer parameter sharing
Parameter Size	110 - 340 million	12 - 235 million	125 - 355 million
Training Efficiency	Slower	Faster	Faster
Performance	Strong performance on diverse tasks	Comparable performance to BERT	Improved performance over BERT
Notable Implementations	BERT, DistilBERT	ALBERT, TinyBERT, ALBERT-xxlarge	RoBERTa-large, RoBERTa-base, RoBERTa-xxlarge

3.5 Model Training and Validation

3.5.1 Model Training

Loading Pre-trained BERT Model and the model weights. These weights are typically trained on a large corpus of text data for a language modeling task. Adding Classification Head specific to the emotion recognition task. This head usually consists of one or more dense layers followed by a Softmax layer for multi-class classification (positive, negative, neutral). This step further includes configuring training parameters such as learning rate, batch size, number of epochs, and optimizer for model training. Iterate over the training dataset for multiple epochs. In each epoch, feed batches of tokenized and padded input sequences along with their corresponding sentiment labels to the model. Compute the forward pass to obtain model predictions. Calculate the loss between the predicted sentiment and the ground truth labels using the chosen loss function. Perform back-propagation to compute gradients and update the model parameters using the chosen optimizer. Monitor training metrics such as loss and accuracy on the training set.

3.5.2 Model Validation

After each training epoch, evaluate the model's performance on the validation dataset. Feed batches of tokenized and padded input sequences from the validation set to the trained model. Compute the forward pass to obtain model predictions. Calculate the validation loss and accuracy. Monitor validation metrics to assess the model's generalization performance and prevent over-fitting.

3.5.3 Hyperparameter Tuning

Perform hyperparameter tuning using techniques like grid search or random search to optimize model performance. Tune parameters such as learning rate, dropout rate, and the number of layers in the classification head. For the implementation of the emotion recognition task, we used Cross-entropy loss function, Adam optimizer, softmax classifier, batch size of 32, and trained for 10 epochs.

3.6 Model Testing and Evaluation

Once training is complete, evaluate the final trained model on the test dataset to assess its performance on unseen data. The model was then evaluated using various performance metrics such as accuracy, precision, recall, and F1-score to assess the model's effectiveness in emotion classification. This step in the implementation of the research is explained in next chapter of the thesis.

CHAPTER 4

EXPERIMENTAL SETUP & RESULT ANALYSIS

In this section, we discuss the experimental setup of the implementation of the models, the libraries used for the task, software programs and hardware components used to run the model and, at last the performance metrics used to assess the model's performances and the overall result analysis of the implementation of the research work.

4.1 Experimental Setup

4.1.1 Software Requirements

1. **Platform**: The models are implemented within the Google Colab environment using Python.
2. **APIs and Drivers**: To ensure smooth operation and compatibility within Google Colab, essential APIs and drivers are installed.
3. **Software**: Google Colab is the chosen software platform due to its accessibility and integration with various Google services.
4. **Language**: Given its extensive use and strong support within the data science community, Python 3.9 was selected as the programming language for this project.

4.1.2 Hardware Requirements

1. **Processing Power**: To run the LLMs we used Tensor Processing Units (TPUs) provided by Google Colab, which offer high-speed computation capabilities required for processing large datasets and complex algorithms.
2. **Memory**: Memory allocation is managed on demand by Google Colab, allowing for flexible scaling based on the model's requirements.
3. **Secondary Storage**: A hard disk or Solid-State Drive (SSD) that meets the requirements of Windows OS 10 is recommended for local storage needs.

4.1.3 Libraries/Packages

i) **Numpy**: NumPy is a library used for scientific computing in Python. It is a powerful tool for numerical operations and data analysis.

ii) **Pandas**: Pandas provides specialized data structures and various tools specifically designed for manipulating numerical tables and time series data.

iii) **Matplotlib**: Matplotlib is the core plotting library for Python. It works seamlessly with NumPy and allows you to create various visualizations, including static charts, interactive plots, and even animations.

iv) **Seaborn:** Seaborn builds on top of Matplotlib, offering a high-level interface specifically tailored for creating statistical graphics. It simplifies the process of generating visually appealing and informative visualizations that effectively communicate insights from data.

v) **Re:** Python's built-in re module provides powerful tools for working with text through regular expressions. These expressions are like search patterns that can identify, extract, and manipulate text data based on specific rules you define.

vi) **String:** The string module is another built-in Python library that offers a variety of functions for working with strings. It covers common operations like combining strings (concatenation), splitting strings into parts, searching for substrings, and modifying string data.

vii) **NLTK:** The Natural Language Toolkit (NLTK) is a popular library specifically designed for natural language processing (NLP) tasks in Python. It provides a comprehensive set of tools for various NLP applications, including splitting text into meaningful units (tokenization), converting words to their root forms (stemming/lemmatization), classifying text by category, and more.

viii) **Demoji:** Demoji is a specialized library designed to handle emojis within text data in Python. It provides functionalities for detecting emojis, replacing them with text descriptions, or even translating them to different languages. This can be extremely useful when working with data that might contain emojis.

ix) **TensorFlow:** TensorFlow is an open-source library widely used for numerical computations, particularly in the realm of deep learning, a powerful subset of machine learning. It allows you to efficiently build and train complex models for various tasks.

x) **Scikit-learn (sklearn):** Scikit-learn (often abbreviated as sklearn) is a popular and versatile Python library for machine learning. It offers a comprehensive set of tools for the entire machine learning workflow, including data preprocessing, classification (grouping data), regression (predicting continuous values), clustering (finding groups in data), and model evaluation. It's known for its user-friendly interface and extensive documentation, making it a great choice for beginners and experts alike.

xi) **Transformers:** Transformers is a powerful open-source library specifically built for Natural Language Processing (NLP) tasks. It provides pre-trained models that have already learned from massive amounts of text data. These models can be fine-tuned for various NLP tasks, such as classifying text by category, answering questions based on text, and summarizing large chunks of text.

This comprehensive setup ensures that the model is well-equipped to handle the computational demands for the emotion recognition task.

4.2 Dataset Description

The dataset used for the implementation of the research work is taken from Kaggle, it consists of tweets, where users shared their opinions and feelings on the COVID-19 outbreak on Twitter. The dataset is carefully designed to include a wide range of tweets from different geolocation, which have variety of linguistic and cultural differences. Fig. 4. shows count of tweets from various locations. The dataset was divided into training, validation, and testing sets after data preprocessing. The distribution of all the datasets is shown in Table 4.1.

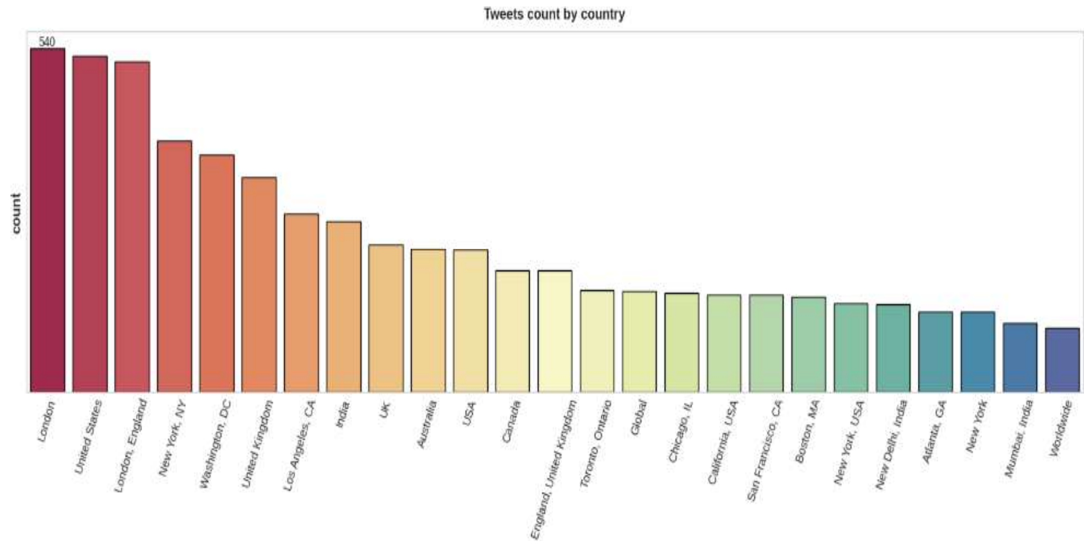


Fig.4.1 Distribution of tweets from different geographical regions

Table 4.1 Distribution of data

Dataset	Count
Training	48597
Validation	5400
Testing	3793

4.3 Performance Evaluation Matrices

Performance evaluation matrices are systematic tools which are used to assess the performance of models. We have used various metrics, such as precision, recall, accuracy, and F1-score in order to evaluate the performance of the models. These metrics are defined as follows:

4.3.1 Confusion Matrix

A confusion matrix is a table that breaks down how well your classification model performed. It shows how many data points were correctly and how many were mistaken for other categories.

i. **Axes:** The confusion matrix has two axes:
   - *Rows*: Represent the actual labels (ground truth) of the data points.
   - *Columns:* Represent the labels predicted by your model.

ii. **Cells:** Each cell in the matrix represents the number of data points that belong to a particular category:

- *True Positives (TP):* These are data points correctly classified as positive by the model. They fall on the diagonal of the matrix (where the row and column labels match).
- *True Negatives (TN):* These are data points correctly classified as negative by the model. They also fall on the diagonal.
- *False Positives (FP):* These are data points incorrectly classified as positive by the model (predicted positive but actually negative).
- *False Negatives (FN):* These are data points incorrectly classified as negative by the model (predicted negative but actually positive).

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	**TP** *True Positive*	**FP** *False Negative*
	Negative	**FP** *False Positive*	**TP** *True Negative*

Table 4.2 Confusion matrix

By analyzing the distribution of values in the confusion matrix, you can gain insights into your model's performance. High values on the diagonal indicate good overall classification accuracy, as the model is correctly identifying most data points. Whereas, High values off the diagonal indicate errors in classification. High FP suggests the model is over-predicting a particular class, while high FN suggests it is under-predicting it.

4.3.2 Precision

Measures the proportion of positive predictions that were actually correct (avoiding false positives). Think of it as how good your model is at being precise with its "positive" labels. It is computed as:

$$\text{Precision} = \frac{TP}{TP+FP}$$

4.3.3 Recall/Sensitivity

Measures the proportion of actual positive cases that were correctly identified (avoiding false negatives). This reflects how well your model captures all the true positive cases. It is calculated as

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

4.3.4 Specificity

This parameter represents the fraction of correctly predicted absence samples relative to the aggregate of absence samples. It is defined as:

$$\text{Specificity}= \frac{TN}{TN+FP}$$

		Predicted Class		
		Positive	**Negative**	
True Class	**Positive**	**TP** *True Positive*	**FP** *False Negative*	**Sensitivity** $\frac{TP}{TP+FN}$
	Negative	**FP** *False Positive*	**TP** *True Negative*	**Specificity** $\frac{TN}{TN+FP}$
		Precision $\frac{TP}{TP+FP}$	**Negative Predicted Value** $\frac{TN}{TN+FN}$	**Accuracy** $\frac{TP+FP}{TP+TN+FP+FN}$

Fig. 4.2 Various Performance Parameters: Confusion Matrix, Precision, Recall, Accuracy, Sensitivity, Specificity, and Negative Predicted Value

4.3.5 Accuracy

Overall percentage of correctly classified data points (both positive and negative). While simple, it can be misleading in imbalanced datasets. It is expressed as:

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

4.3.6 F1 Score

F1 Score combines precision and recall into a single metric, providing a balanced view of model performance. It penalizes models that excel in one metric but struggle in the other. It is calculated as:

$$\text{F1 Score} = \frac{2*(Precision*Reall)}{(\text{Precision} + \text{Recall})}$$

4.4 RESULT ANALYSIS

Initially, the dataset was subjected to preprocessing steps, including tokenization, lowercasing, and data cleaning, which included handling null values, and categorical values, and removing outliers, hashtags, links, mentions, emoji, and special characters.

Pre-trained models were imported from the HuggingFace transformer library during the model-development phase. These models were then trained using the provided dataset, incorporating a fine-tuning approach whereby the Adam optimizer and Softmax classifier were employed. The models' performance was assessed using the testing dataset using performance metrics like accuracy, precision, recall, F1-score, classification report, and confusion matrix. Fig presents the confusion matrix for all the models implemented in chapter 4. Table 4 presents the results for emotion recognition task using Naïve Bayes, BERT, ROBERTA, ALBERT models, which show that transformer-based models perform better than conventional machine learning (ML) methods.

In particular, RoBERTa outperforms BERT and ALBERT in terms of F1 scores among the transformer models evaluated. These results demonstrate the effectiveness LLMs, in particular RoBERTa.

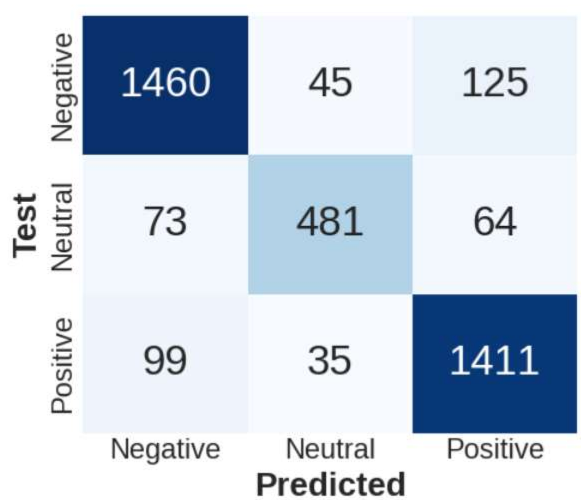


Fig. 4.3 Confusion Matrix for emotion recognition using BERT

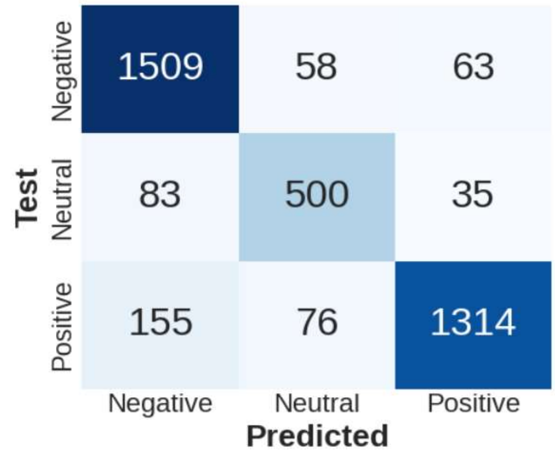


Fig. 4.4 Confusion Matrix for emotion recognition using ROBERTA

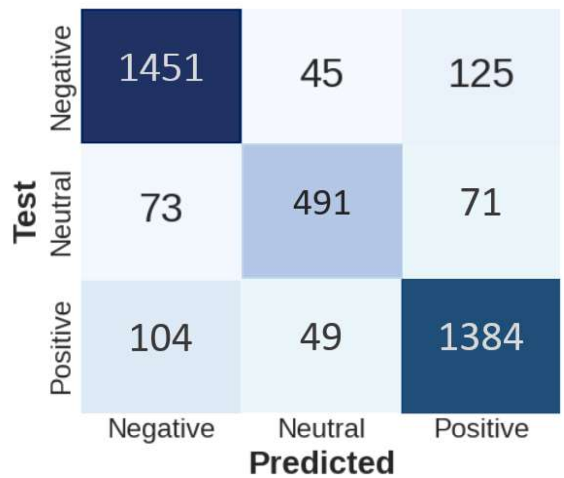


Fig.4.5 Confusion Matrix for emotion recognition using ALBERT

Table 4.3 Performance evaluation of the models

Model	Label	Precision	Recall	F1-score
Naïve Bayes	Negative	0.7	0.78	0.74
	Neutral	0.6	0.47	0.53
	Positive	0.73	0.72	0.72
BERT	Negative	0.88	0.91	0.89
	Neutral	0.89	0.75	0.81
	Positive	0.89	0.91	0.90
RoBERTa	Negative	0.91	0.92	0.91
	Neutral	0.74	0.84	0.79
	Positive	0.92	0.92	0.92
ALBERT	Negative	0.86	0.89	0.87
	Neutral	0.89	0.78	0.83
	Positive	0.86	0.89	0.87

CHAPTER 5

CONCLUSION, FUTURE WORKAND CHALLENGES

This section discusses the overall conclusion of the research work carried out in this thesis, and outlines the challenges and future work for the emotion recognition task utilizing LLMs.

5.1 Conclusion

In this thesis, we explored the potential of Large Language Models (LLMs) for emotion recognition through sentiment analysis. It provides an analysis of various fine-tuned LLMs across different domains and languages. It critically examines how different LLMs, such as BERT, GPT-3, and other Transformer-based models, are adapted and fine-tuned for specific emotional recognition tasks We began with a comprehensive literature survey, examining existing emotion recognition techniques and the role of LLMs in sentiment analysis. The survey in this study highlights the effectiveness of the models in diverse sectors, including social media analysis, customer service, and mental health assessment, showcasing their versatility in processing and interpreting emotional content. Additionally, the survey delves into the performance of LLMs on datasets in multiple languages, underscoring the challenges and successes in capturing the nuances of emotional expression in non-English languages [42]. Notably, LLMs exhibit an outstanding ability for processing and understanding emotional expressions in a variety of languages. The efficacy of LLMs over other traditional models is due to their extensive pre-training on vast corpora covering a broad spectrum of linguistic and cultural nuances, combined with advanced techniques for fine-tuning that customize performance in recognizing emotions, these strategies allow the models to fine-tune their understanding of emotional expressions within particular linguistic and cultural context the models for particular emotional recognition tasks and datasets[43,44]. The survey also highlights various innovative approaches for fine-tuning that have been devised to enhance LLM, offering insights into the models' capabilities and limitations in global applications.

Following the survey, we conducted an experiment employing various LLM architectures for sentiment analysis tasks. The experiment analyzed the effectiveness of these models in identifying emotions within textual data. This study explored the effectiveness of utilizing various transformer models, specifically BERT, RoBERTa, and ALBERT, for the emotion recognition task. Through extensive experimentation and evaluation, the paper showcased the remarkable abilities of these models to recognize and categorize human emotions from a variety of textual inputs. The findings of this paper highlight the potential of leveraging self-attention mechanisms and pre-trained representations to achieve state-of-the-art performance in emotion recognition tasks. Moreover, by contrasting the performance and computational efficiency of each model, this work provided significant findings regarding the comparative benefits and compromises of every model concerning emotion identification. Pre-trained transformer models, such as ALBERT, RoBERTa, and BERT, are important developments in language processing that allow machines to understand as well as interact with human language in a variety of ways.

Furthermore, the thesis discusses the evolving nature of emotion recognition technology, including ethical considerations and future prospects. It suggests areas for further research, emphasizing the need for more inclusive and representative datasets and the exploration of multimodal emotion recognition approaches [45]. This thesis serves as a valuable resource for understanding the current state and future potential of emotion recognition using fine-tuned LLMs in a global context.

5.2 Challenges and Future work

Large Language Models (LLMs) based emotion recognition is a burgeoning area that offers an incredible number of opportunities, problems, and new directions for further research. Although LLMs have transformed our capacity to understand and evaluate human emotions in written communication, they also introduce challenges that need to be appropriately addressed [47,48]. This study explores many challenges, opportunities and future scope in the field of emotion recognition using LLM which are mentioned below sections.

5.2.1 Challenges

1. **Privacy:** A key challenge for models is ensuring data privacy. These models often require access to extensive datasets, potentially containing sensitive information, to make accurate predictions. This raises significant concerns about data confidentiality and the risks associated with improper data use or management. Balancing the benefits of models with the need to protect user privacy is a critical issue, necessitating robust privacy safeguards and clear data management policies.

2. **Social Bias:** Large amounts of text data are used to train LLMs, which may unintentionally introduce the existing biases from the training set into the model's results. Such biases can perpetuate and amplify societal prejudices, leading to unfair or skewed results. Addressing social bias in is essential for promoting fairness, inclusivity, and unbiased decision-making. Ongoing efforts include incorporating diverse and representative datasets to reduce bias transmission and developing techniques for identifying and mitigating bias during the model's training phase.

3. **Computational and Financial Costs:** Deploying GPT models involves substantial computational resources, including extensive processing time, data storage requirements, and energy consumption. This makes them computationally intensive and financially demanding, posing a challenge for their widespread application and development. Enhancing the efficiency and cost-effectiveness of these models remains a crucial area of focus in the field.

4. **Interpreting lengthy texts:** These models are effective at identifying patterns in the data that they have been trained on, but they may not perform as well in complex scenarios requiring in-depth understanding across lengthy text passages. This restriction may lead to inaccuracies in tasks involving complex comprehension. This difficulty highlights the distinction between recognizing linguistic patterns and understanding or reasoning like a human, particularly in circumstances that need a sophisticated understanding of the subject matter throughout a larger story.

5. **Insufficient training datasets for specialized domains/language:** Nevertheless, their effectiveness may be limited when utilized in highly specialized domains or languages that lack sufficient training datasets. This is due to the possibility that these models, even with their intensive training, do not have the specialized vocabulary, knowledge of particular terminologies, or subtleties particular to certain domains. Their efficiency at tasks involving domain-specific knowledge or subtle linguistic differences may be impacted by this constraint. To address this limitation, domain-specific fine-tuning can be done where the models are fine-tuned on a corpus of text that is specific to the domain in question, this training dataset can be enhanced using techniques like data augmentation to create additional synthetic samples relevant to the domain. Also, new models can be developed that are extensively pre-trained on domain-specific corpora, which ensures the alignment of the model's foundational knowledge with the requirements of the domain or language.

5.2.2 Future Work

1. **Enhancing Model Robustness:** Subsequent investigations ought to put emphasis on enhancing the robustness of large language models (LLMs) to accommodate diverse data quality and contextual variations. Improving LLMs' accuracy in deciphering complex or ambiguous emotional signals is the objective. This involves creating models that can effectively understand complicated and nuanced emotional cues, which are frequently present in natural language, in addition to being skilled with obvious emotional indicators.
2. **Cross-Cultural and Multilingual Studies:** Extending the research's reach to include a larger range of languages and cultures is another important subject for future research. It is essential to comprehend how different linguistic and cultural backgrounds portray emotions in different ways. More inclusive and globally applicable emotion recognition systems that are mindful of cultural variations in emotional expression are required.
3. **Bias Considerations in LLM:** It is crucial to address at potential biases in

LLM-based emotion recognition and to establish standards to address them. This assures that technology is used ethically by taking privacy concerns into account and attempting to lessen biases that can result from the model's design or training set. Creating these standards will aid in the development of more equitable and accountable AI systems.

4. **Integration with Multimodal Data:** Subsequent research ought to investigate the incorporation of text-based emotion recognition with different data formats, like auditory and visual inputs. This multimodal method, which integrates verbal cues with non-verbal ones like tone of voice or facial expressions, can result in a more detailed understanding of emotions by offering a richer and more comprehensive picture of emotional states.
5. **Real-time Processing and Applications:** For practical applications, techniques for quick, real-time emotion recognition must be developed. This covers applications such as monitoring mental health, improving interactions with customers, and assisting with instructional programs. In domains where instantaneous feedback or contact is critical, real-time emotion recognition can greatly help, opening the door for more adaptable and responsive artificial intelligence systems.

5.3 Social and Industrial Application

The advancements made in emotion recognition using LLMs, as explored in this thesis, hold promise for various applications across social and industrial sectors. Here are some potential areas where this technology could be implemented:

5.3.1 Social Applications:

1. **Mental Health Support:** LLMs can analyze text from chat bots or online support groups to identify users expressing negative emotions. This could enable early intervention and referral to mental health resources.
2. **Personalized Education:** Educational platforms can leverage LLM-based emotion recognition to gauge student engagement and adjust teaching styles based on their emotional state.
3. **Enhanced Social Media Interactions:** Social media platforms could utilize LLMs to detect cyberbullying or hateful content, fostering safer

online communities.

4. Improved Customer Service: Chatbots powered by LLMs can analyze customer sentiment during interactions, allowing for more empathetic and personalized service.

5.3.2 Industrial Applications:

1. **Market Research:** Businesses can analyze customer reviews and social media posts using LLMs to understand emotional responses to products and services, informing marketing strategies.
2. **Human Resource Management:** LLMs can be used to analyze employee sentiment during surveys or exit interviews, providing valuable insights into company culture and employee satisfaction.
3. **Product Design and Development:** LLMs can analyze user feedback on prototypes or product designs, helping companies identify emotional responses and refine products accordingly.
4. **Personalized Advertising:** Advertising platforms can leverage LLM-based emotion recognition to tailor ad content based on a user's emotional state, potentially increasing engagement and effectiveness.

Overall, the research on emotion recognition using LLMs offers a powerful tool for understanding human emotions within various social and industrial contexts. As the technology matures, its applications have the potential to revolutionize how we interact with each other and with technology itself.

REFERENCES

1. Shanahan, M. (2022). Talking about large language models. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2212.03551

2. Hadi, M. U., Tashi, Q. A., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., & Mirjalili, S. (2023). Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects. Research Gate. https://doi.org/10.36227/techrxiv.23589741.v3

3. Miyazawa K, Kyuragi Y, Nagai T Simple and effective multimodal learning based on Pre-Trained Transformer models. IEEE Access 10:29821–29833. https://doi.org/10.1109/access.2022.3159346 (2022)

4. Zhang B, Yang H, Zhou T, Babar A, Liu X: Enhancing financial sentiment analysis via retrieval augmented large language models. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2310.04027 (2023)

5. Patel A, Oza P, Agrawal S: Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model. Procedia Computer Science 218:2459–2467. https://doi.org/10.1016/j.procs.2023.01.221 (2023)

6. A. Mallik, S. Kumar, "Word2Vec and LSTM based deep learning technique for context-free fake news detection". Multimedia Tools and Applications, 83, 919–940, (2024)2. A. Kumar, D.K. Jain, A. Mallik, S. Kumar, "Modified node2vec and attention based fusion framework for next POI recommendation". Information Fusion, 101, 101998 (2024)

7. S. Kumar, A. Mallik, " COVID-19 Detection from Chest X-rays Using Trained Output Based Transfer Learning Approach". Neural processing letters, 55(3), 2405-2428 (2023)4. S. Kumar, A Mallik, S.S Sengar, "Community detection in complex networks using stacked autoencoders and crow search algorithm". The Journal of Supercomputing, 79(3), 3329-3356 (2023)

8. A. Bhowmik, S. Kumar, N. Bhat, "Evolution of automatic visual description techniques-a methodological survey". Multimedia Tools and Applications, 80(18), 28015-28059 (2021)

9. D. Kurchaniya, S. Kumar, "Two stream deep neural network based framework to detect abnormal human activities", Journal of Electronic Imaging 32 (4), 043021-043021 (2023)

10. E. Mahajan, H. Mahajan, S. Kumar, "EnsMulHateCyb: Multilingual hate

speech and cyberbully detection in online social media", Expert Systems with Applications 236, 121228, (2024)

11. Deng, J., & Ren, F. (2023). A survey of textual emotion recognition and its challenges. IEEE Transactions on Affective Computing, 14(1), 49–67. https://doi.org/10.1109/taffc.2021.3053275

12. Ray, P. P.: ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, 3, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003 (2023)

13. Yang, D., Kommineni, A., Alshehri, M., Mohanty, N., Modi, V., Gratch, J., & Narayanan, S. (2023). Context Unlocks Emotions: Text-based Emotion Classification Dataset Auditing with Large Language Models. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2311.03551

14. N. Kolt, "Predicting consumer contracts," Berkeley Tech. LJ, vol. 37, p. 71, 2022.

15. Cohen, D. Presil, et al.: Enhancing social network hate detection using back translation and gpt-3 augmentations during training and test-time, Information Fusion, p. 101887, (2023).

16. L. Chiu, A. Collins, and R. Alexander, Detecting hate speech with gpt-3, arXiv preprint arXiv:2103.12407, (2021).

17. T. Lin, Y. Wang, X. Liu, and X. Qiu, "A Survey of Transformers," Jun. 2021, http://arxiv.org/abs/2106.04554

18. A. Vaswani et al., "Attention Is All You Need," Jun. 2017, http://arxiv.org/abs/1706.03762

19. Deng X, Bashlovkina V, Han F, Baumgartner S, Bendersky M :LLMs to the Moon? Reddit Market Sentiment Analysis with Large Language Models. ACM Web Conference. https://doi.org/10.1145/3543873.3587605 (2023)

20. Mao R, Liu Q, He K, Li W, Wang Z: The Biases of Pre-Trained Language Models: An Empirical study on Prompt-Based Sentiment Analysis and Emotion Detection. IEEE Transactions on Affective Computing 14(3):1743–1753. https://doi.org/10.1109/taffc.2022.3204972 (2023)

21. Güven, Z. A.: The Comparison of Language Models with a Novel Text Filtering Approach for Turkish Sentiment Analysis. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(2), 1–16. https://doi.org/10.1145/3557892 (2022)

22. Üveges I, Ring O: HunEmBERT: A Fine-Tuned BERT-Model for classifying sentiment and emotion in political communication. IEEE Access 11:60267–60278. https://doi.org/10.1109/access.2023.3285536 (2023)

23. Roumeliotis KI, Tselikas ND, Nasiopoulos DK: LLMs in e-commerce: A comparative analysis of GPT and LLaMA models in product review evaluation. Natural Language Processing Journal :100056. https://doi.org/10.1016/j.nlp.2024.100056 (2024)

24. Mohammad SM, Bravo-Márquez F: WASSA-2017 Shared Task on Emotion Intensity. arXiv (Cornell University) (2017)

25. Chowdhary, K., et al.: Natural language processing. Fundamentals of artificial intelligence, pp.603-649 (2020)

26. Thakur, P., D.R. Shrivastava, et al., A: A review on text-based emotion recognition system. International Journal of Advanced Trends in Computer Science and Engineering, 7(5) (2018)

27. Devlin, J. M.W. Chang, K. Lee, and K. Toutanova: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

28. Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma and Soricut, R.: Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942 (2019)

29. Tan, K.L., Lee, C.P., Anbananthen, K.S.M. and Lim, K.M.: RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network. IEEE Access, 10, pp.21517-21525 (2022)

30. Sanh, V., Debut, L., Chaumond, J. and Wolf, T.: DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108 (2019)

31. Deng, J. and Ren, F.: A survey of textual emotion recognition and its challenges. IEEE Transactions on Affective Computing (2021)

32. Yadav, N., Kudale, O., Rao, A., Gupta, S. and Shitole, A.: Twitter sentiment analysis using supervised machine learning. In Intelligent data communication technologies and internet of things: Proceedings of ICICI 2020 (pp. 631-642). Springer Singapore (2021)

33. Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L.T. and Trajanov, D.: Evaluation of sentiment analysis in finance: from lexicons to transformers. IEEE access, 8, pp.131662- 131682 (2020)

34. Deng, J. and Ren, F.: A survey of textual emotion recognition and its challenges. IEEE Transactions on Affective Computing (2021)

35. RoBERTa.(n.d.). https://huggingface.co/docs/transformers/model_doc/roberta (2024)

36. ALBERT.(n.d.). https://huggingface.co/docs/transformers/model_doc/albert (2024)

37. BERT.(n.d.). https://huggingface.co/docs/transformers/model_doc/bert (2024)

38. Bikku, T., Jarugula, J., Kongala, L., Tummala, N.D. and Donthiboina, N.V.: June. Exploring the effectiveness of BERT for sentiment analysis on large-scale social media data. In 2023 3rd International Conference on Intelligent Technologies (CONIT) (pp. 1-4). IEEE (2023)

39. Prasanthi, K.N., Madhavi, R.E., Sabarinadh, D.N.S. and Sravani, B.: April. A Novel Approach for Sentiment Analysis on social media using BERT & ROBERTA Transformer-Based Models. In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE (2023)

40. Shlkamy, E.S.I., Mahar, K.M. and Sedky, A.A.H. : A Russia-Ukraine Conflict Tweets Senti- ment Analysis Using Bidirectional LSTM Network (2023)

41. Sudhir, P. and Suresh, V.D.: Comparative study of various approaches, applications, and classifiers for sentiment analysis. Global Transitions Proceedings, 2(2), pp.205-211 (2021)

42. Bozuyla, M.: Sentiment Analysis of Turkish Drug Reviews with Bidirectional Encoder Representations from Transformers. ACM Transactions on Asian and Low-Resource Language Information Processing, 23(1), pp.1-17 (2024)

43. Gupta, P., Gandhi, S. and Chakravarthi, B.R.: December. Leveraging transfer learning techniques-bert, Roberta, Albert, and distilbert for fake review detection. In Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation (pp. 75-82) (2021)

44. Dataset. https://www.kaggle.com/code/ (2024)

45. Goel, A., Gautam, J. and Kumar, S.: October. Real-time sentiment analysis of tweets using Naive Bayes. In 2016 2nd International Conference on Next Generation Computing Technologies (NGCT) (pp. 257-261). IEEE (2016)

46. Han, X., Wang, Y.T., Feng, J.L., Deng, C., Chen, Z.H., Huang, Y.A., Su, H., Hu, L. and Hu, P.W.: A survey of transformer-based multimodal pre-trained models. Neurocomputing, 515, pp.89-106 (2023)

47. Kenton, Jacob Devlin Ming-Wei Chang, and Lee Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding: " Proceedings of naacL-HLT. Vol. (2019)

48. Singla, S. and Ramachandra, N.: Comparative analysis of transformer-based pre-trained NLP Models. Int. J. Comput. Sci. Eng, 8, pp.40-44 (2020)

49. Sun, L., Xu, M., Lian, Z., Liu, B., Tao, J., Wang, M., and Cheng, Y.: Multimodal emotion recognition and sentiment analysis via attention enhanced recurrent model. In Proceedings of the 2nd on Multimodal Sentiment Analysis Challenge (pp. 15-20) (2021)

LIST OF PUBLICATIONS/ACCEPTANCES & THEIR PROOFS

1. Vershika Srivastava and Sanjay Kumar, "**Emotion Recognition Using Large Language Models: An Overview**" accepted and presented in '3rd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (IEEE ICPEICES-2024)' at DTU in April 2024.

Certificate of Participation:

IEEE

IEEE ICPEICES-2024

3rd IEEE INTERNATIONAL CONFERENCE

on

POWER ELECTRONICS, INTELLIGENT CONTROL AND ENERGY SYSTEMS

26th-28th APRIL, 2024

Department of Electrical Engineering, Delhi Technological University

This certificate of Participation/Presentation is awarded to

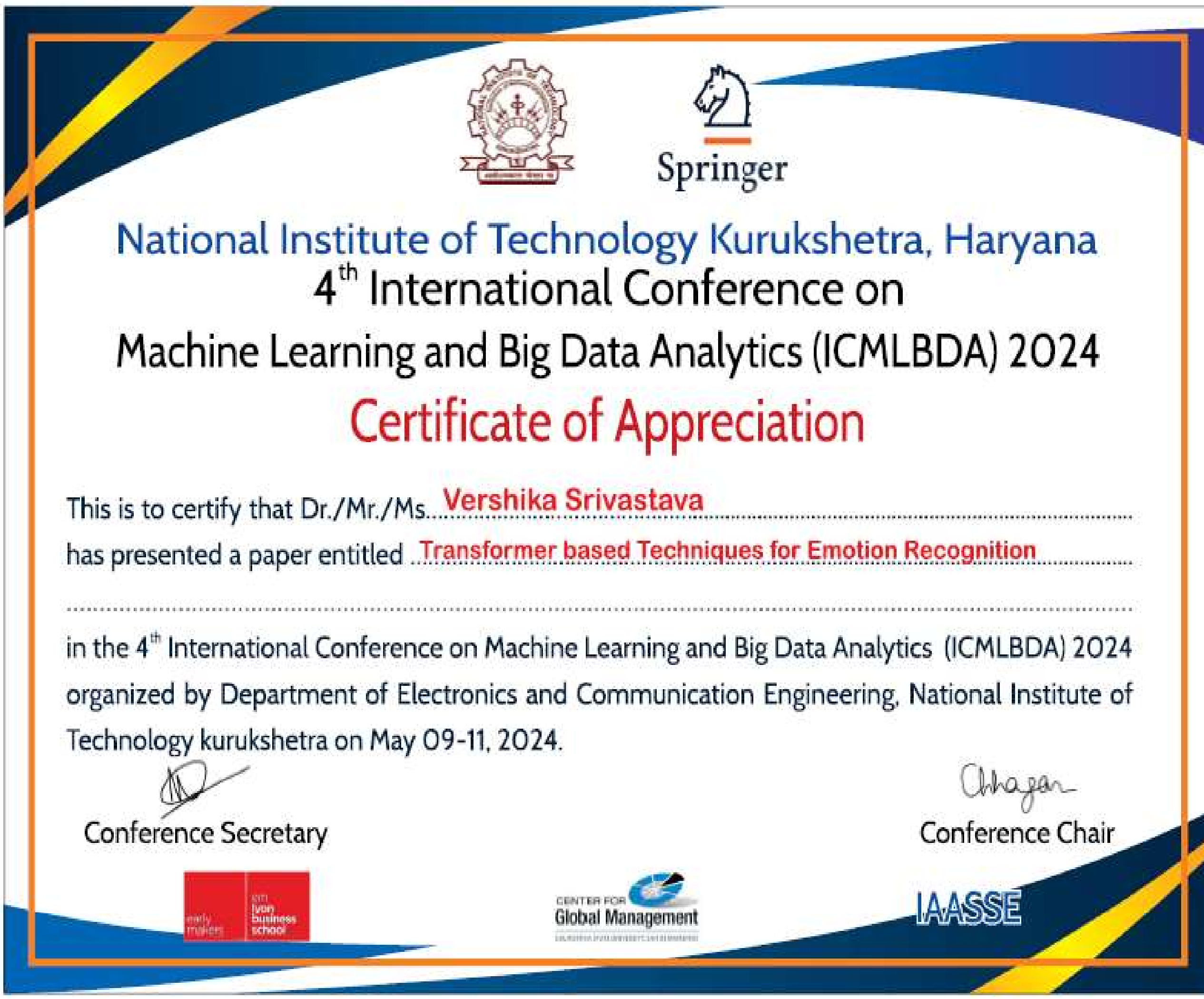
Dr./Mr./Ms. **Vershika Srivastava** has attended the conference and presented paper entitled **Emotion Recognition Using Large Language Models: An Overview** authored by **Vershika Srivastava; Sanjay Kumar** in the **3rd IEEE International Conference on Power Electronics, Intelligent Control** and Energy Systems held at Delhi Technological University, Delhi, India from 26th - 28th April, 2024.

Prof. Vinod Kumar Yadav
General Chair

Prof. Rachana Garg
General Chair

2. Vershika Srivastava and Sanjay Kumar, "**Transformer Based Techniques for Emotion Recognition**" was accepted and presented at '4th International Conference on Machine Learning and Big Data Analytics (ICMLBDA-2024)' at NIT Kurukshetra in May 2024

Certificate of Participation:


Springer

National Institute of Technology Kurukshetra, Haryana

4th International Conference on

Machine Learning and Big Data Analytics (ICMLBDA) 2024

Certificate of Appreciation

This is to certify that Dr./Mr./Ms. Vershika Srivastava

has presented a paper entitled Transformer based Techniques for Emotion Recognition

in the 4th International Conference on Machine Learning and Big Data Analytics (ICMLBDA) 2024 organized by Department of Electronics and Communication Engineering, National Institute of Technology kurukshetra on May 09-11, 2024.

Conference Secretary

Conference Chair

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis: **Emotion Recognition using Large Language Model**

Total Pages **57**

Name of the Scholar: **Vershika Srivastava, (2K22/AFI/28)**

Supervisor: **Dr. Sanjay Kumar**

Department: **Department of Computer Science and Engineering**

This is to report that the above thesis was scanned for similarity detection and outcomes are given below:

Software used: **Turnitin** Similarity Index: 14 **%** Total Word Count: 12425

Date: 5/31/24

Candidate's Signature **Signature of Supervisor**

PAPER NAME

vershika thesis

WORD COUNT	CHARACTER COUNT
12425 Words	71418 Characters
PAGE COUNT	FILE SIZE
57 Pages	1.5MB
SUBMISSION DATE	REPORT DATE
May 31, 2024 12:52 PM GMT+5:30	May 31, 2024 12:53 PM GMT+5:30

14% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 9% Internet database
- 6% Publications database
- Crossref database
- Crossref Posted Content database
- 10% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 10 words)