Journal of Information Systems Engineering and Management
2025, 10(23s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Comprehensive Analysis of the Current Methods, Challenges and Innovations in Sentiment Analysis

[1]Shivani Awasthi, [2]Inderpreet Kaur, [3]Mandeep Kaur
[1]Research Scholar, Computer Science & Engineering Department, SSCSE, Sharda University, Greater Noida, India & Assistant Professor, CSE–AI&ML Department, ABES Engineering College, Ghaziabad, India
shivanigarg1804@gmail.com
shivani.garg@abes.ac.in
[2]Associate Professor, Computer Science & Engineering Department, SSCSE, Sharda University, Greater Noida, India
kaur.lamba1@gmail.com
inderpreet.kaur@sharda.ac.in
[3]Professor, School of Computer Science Engineering & Technology Bennett University, Greater Noida, Uttar Pradesh, India
mandeep.kaur@bennett.edu.in
drmandeepkaur10@gmail.com

ARTICLE INFO

Received: 14 Dec 2024

Revised: 30 Jan 2025

Accepted: 18 Feb 2025

ABSTRACT

Sentiment analysis, also referred to as opinion mining, is an important segment of natural language processing (NLP) and machine learning that textual data targets based on the identification or classification of emotions, opinions, sentiments, or attitudes. Initial works associated with the study of public opinion and the subjectivity of text date from the beginning of the 20th century, while sentiment analysis started to take flight in the late 1990s with the upsurge of user-generated content on the internet. While early approaches were lexicon-based, thereby based on predefined word lists, lately, these have evolved to include machine learning models such as Naive Bayes, Support Vector Machines (SVMs), and deep learning techniques like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Of late, research has focused more on fine-grained or subtle sentiment analysis that faces challenges like sarcasm, context-based meaning, and generalization across domains. It has further evolved into multimodal analysis, combining text with other forms of media, such as images and audio. In spite of this, all challenges remain in cross-lingual sentiment detection and ethical considerations within sentiment analysis. Future research will probably be directed at enhancing the accuracy of the models, real-time processing, and addressing language diversity in the application of sentiment analysis.

Keywords: Sentiment analysis, Natural Language Processing, Convolutional Neural Networks, Recurrent Neural Networks, Support Vector Machines

1. INTRODUCTION

Sentiment analysis, over the recent years, has gained such momentum that it has today become one of the primary areas of research in NLP and machine-related learning. It is a method of textual data analysis that identifies feelings, opinions, and attitudes usually positive, negative, or neutral. A lot of work is being conducted today based on sentiment analysis, especially in various domains like product reviews, social media platforms, and even political discourses. It helps in business with a clearer understanding of feedback from customers, and thus it is a very strong tool in the analysis of market trends. Recent studies into sentiment analysis have emphasized precision based on methods of advanced machine learning and deep learning models, thus holding great promise in dealing with large volumes of unstructured data.

The ever-increasing demand for accurate sentiment analysis has, therefore, resorted to the application of neural networks, especially CNN and other deep learning approaches. These boast results that are much more robust than the ones obtained with traditional approaches. For example, some of the CNN models did an excellent job in classifying movie reviews as positive or negative with performance metrics as high as 88.96% for precision, depicted by Lou (2023). Other recent works employed machine learning algorithms, such as SVMs and logistic regression to enhance the model for higher precision and recall of aspects. For instance, the recent work of Do et al., 2019, uses these techniques to perform aspect-based sentiment detection.

As sentiment analysis is continuously evolving, more and more research focuses on deeper analysis that surpass the simple positive/negative distinction in order to capture nuanced opinions that can give more insight into consumer behaviour and public sentiment. So far, this all shows the continuously improved importance of the field and great scope for improvement in the coming years both in the accuracy of results and regarding applications.

Historically, traditional sentiment analysis techniques were posed to be mainly unimodal, focused mostly on the textual data in terms of extracting the sentiments. However, the growth of user-generated multimedia content on social platforms has brought multimodal sentiment analysis, which likely integrates text, audio, visual, and contextual information, much into play, promising much more nuance and accuracy of understanding of sentiments (Cambria et al., 2017). From unimodal to multimodal techniques, this approach reflects the increasing realization that human communication is intrinsically multimodal, with expressions of emotion often conveyed through language, tone, facial expressions, and gestures (Soleymani et al., 2017).

Unimodal techniques for sentiment analysis are mostly limited to just one type of data, which is always textual in nature and depends on various methods, such as from lexicon-based models to deep learning architectures with their extensions into recurrent neural networks and transformers. Despite impressive advances in text-based sentiment analysis, such unimodal approaches fail to capture the intricacies of complex emotional expression when textual cues might be inconclusive or incomplete (Zadeh et al., 2018). For example, sarcasm and irony are often of such a kind that they need extralinguistic cues - if you want to talk about things like tone of voice or facial expression - that cannot be acquired from text.

The issues discussed above are addressed by multimodal sentiment analysis by combining multiple streams of data to be able to make an overall and more holistic view in understanding sentiment. For instance, the promising results that were obtained from the fusion of textual, visual, and acoustic multimodality approaches since each modality may fill in the weaknesses of others (Poria et al., 2019). Multimodal techniques usually rely on sophisticated machine learning models that deal with heterogeneous data, such as multimodal transformers, tailored convolutional neural networks (CNNs) for image data and audio processing. Moreover, the combination of multi-source data tends to cause multimodal techniques to perform better than unimodal models in terms of precision and robustness, especially in the context of heavy complexities involving social media, where users tend to freely intermix words, images, and other multimedia elements to convey their feelings (Maiya & Patil, 2020).

This paper explores and critically analyzes the differences, as well as strengths and weaknesses, of unimodal and multimodal techniques as relates to sentiment analysis. The comparative analysis used here is aimed at throwing into sharp relief the progress made in multimodal approaches and highlighting the challenges and nuances in model deployment in various contexts. As part of our contribution, we also offer perspectives on future research directions in this fast-evolving area of study, particularly highlighting the potential of improved sentiment analysis capabilities for applications such as brand monitoring and customer feedback systems as well as human-computer interaction.

This paper is organised as follows: we begin with the outline of the motivation behind carrying out this work of comprehensive and critical investigation of sentiment analysis. We, then, discuss the progression of sentiment analysis over the recent years followed by its key objectives and reach. Next, we explore the methodologies and techniques used in sentiment analysis, including both classical and emerging techniques, placing the spotlight on specific methods and techniques widely used in practice. With this background, we then delve a little deeper into certain real-world applications of this field followed by the datasets used for building and testing sentiment analysis models along with the evaluation metrics used to assess the efficiency of these models. Our study comes towards conclusion after discussing the recent challenges faced in this field and how use of multiple modalities can provide a major boost to the efficiency of sentiment analysis techniques.

2. MOTIVATION

The motivation towards the detailed examination of sentiment analysis arises due to the rapid growth and diverse applications in several domains, ranging from product reviews to social media and financial markets. It has evolved from simple models of text classification into more sophisticated ones—the current state of the art in attempts to capture subtle emotions and opinions from large volumes of unstructured data. However, human language complexity, including context, sarcasm, and cultural variations, points to major barriers in the accuracy and generalization of sentiment analysis models. This paper therefore calls for a critical review of the existing methodologies, identifying the gaps that exist in existing research, and suggesting areas which could be potentially improved upon.

With businesses increasingly dependent on sentiment analysis as an important strategic input to business decision making, the demand is increasing for more accurate models that rightly portray consumer sentiment on each platform. Traditional methods, such as the lexicon-based approach, fail to handle complex linguistic features like negation and irony; there has always been an additional motive, therefore, to provide a general overview in this field (Chavan et al., 2018). Again, the development of multimodal sentiment analysis, which combines text with images and audio, is a potential and promising avenue. This aspect of sentiment analysis is still underexplored—the review will make clear how such modalities may extend traditional text-based methods. This would, finally, provide a critical roadmap for advancing sentiment analysis research and making it useful in practice (Mäntylä et al., 2016).

3. THE PROGRESSION OF SENTIMENT ANALYSIS

Sentiment analysis actually has its origins in the early 20th century, when public opinion research and the study of emotions within psychology and sociology began to look at human sentiments and their impact on behaviour. Still, it wasn't until much later, during the late 1990s and the early 2000s, that sentiment analysis in computational linguistics and NLP really took off, as the internet and user-generated content blew through the roof of critical mass.

The first significant move toward sentiment analysis was done by the studies of text subjectivity analysis, which had the aim of observing subjective opinions in texts rather than objective facts, conducted by (Mäntylä et al., 2016). Early Sentiment Analysis was thus primarily lexicon-based, relying on pre-defined lists of words labeled with their emotional values. With the increase in subjective text on the web, especially through social networking sites, the focus of sentiment analysis also began to extend to new domains such as product reviews, political discourse, and further to social media platforms like Twitter and Facebook.

Until the early 2000s, machine learning-based approaches re-emerged where algorithms are trained from a large dataset of labeled opinions. These methods rectified some of the obvious deficiencies of lexicon-based methods in handling negation and context-specific meaning. But the real breakthrough came with more sophisticated models, such as Naive Bayes, SVMs, and later on deep learning techniques like CNNs or even RNNs.

Sentiment analysis has gradually started to expand from simple polarity detection, such as positive, negative, and neutral, toward more fine-grained tasks, which include aspect-based sentiment analysis targeting at identifying sentiments about particular specific aspects related to the product or service. This more refined analysis has also proved useful in applications such as customer feedback and financial markets, where sentiments often relate to multiple attributes in the same text.

However, sentiment analysis only started to grow steeply after the year 2000 with the advancement of Web 2.0 and the sudden increase in subjective content from user contributions on blogs, reviews, and social media. The computational treatment of such content demanded opinion- and emotion-classification techniques from large datasets, thus motivating substantial research into machine learning methods for sentiment analysis (Mäntylä et al., 2016). Early approaches in sentiment-analysis development included lexicon-based methods developed using pre-defined lists of words tagged with their respective sentiment classification. However, these have limited capabilities regarding nuances like sarcasm or context-specific meanings.

Further research brought the development of more advanced techniques with the help of machine learning, which improved the accuracy of sentiment classification. By the mid-2010s, the integration of deep learning algorithms—enabled by quite remarkable CNNs and Long Short-Term Memory (LSTMs)—permitted the elicitation of both semantic and syntactic aspects of language for better performance. Despite such advances, challenges still remain in handling sarcasm, irony, and context-based sentiments, which are far more complex than capturing simple polarity. Notice that the expansion of sentiment analysis to languages other than English has progressed with the gradual introduction of new variables that furthered another barrier, cross-lingual aspect, in sentiment analysis. Researchers are still struggling with language-specific nuances and a lack of enough labeled multilingual data sets.

4. KEY OBJECTIVES AND REACH OF SENTIMENT ANALYSIS

The scope of this comprehensive analysis on Sentiment Analysis is to review the diversity of techniques in sentiment analysis, their applications across various domains, and further challenges in the field. In recent times, Sentiment Analysis has emerged as a subfield of natural language processing, receiving considerable attention in academe since it is capable of automatically extracting subjective information from the large volume of unstructured data. The review will cover sentiment classification methods using machine learning and deep learning models, emerging trends in aspect-based sentiment analysis, and multimodal sentiment analysis. Another objective of the review is to present how different methods perform in various domains: social media, product reviews, and financial markets.

The objectives of this analytical assessment are as follows:

1. Analyse Existing Approaches: The review will consider the different techniques in sentiment analysis, from the classic lexicon-based ones to the machine and deep learning methods popularly proposed nowadays, and their respective abilities to handle intricate datasets (Mäntylä et al., 2016).
2. Domain-Specific Applications: The review, therefore, focuses on the domain-based uses of sentiment analysis in social media, stock market predictions, and product reviews to show how the sentiments extracted apply in decision-making processes.
3. Limitations and Challenges: The review discusses various limitations of existing models on sentiment analysis, especially with regard to meeting the challenges of sarcasm, extracting context-based sentiment, and multimodal sentiment analysis.
4. Propose Future Research Directions: The review will highlight gaps in current research and propose future directions, such as improving model generalization across domains and exploring hybrid models that combine rule-based and machine learning approaches.

With these objectives in mind, the review tries to give a complete overview of sentiment analysis whose further development is possible in the direction of granularity and accuracy of sentiment detection models in various contexts.

5. SENTIMENT ANALYSIS METHODS AND TECHNIQUES

Sentiment analysis has now become one of the key domains of NLP and machine learning; it has covered a number of methods and approaches. These approaches are meant for opinion/sentiment extraction, classification, and estimation from texts. Following are the major methodologies involved with sentiment analysis:

- **Lexicon-based approaches:** Lexicon-based approaches in sentiment analysis depend on predefined lists called sentiment lexicons. In these, every word has a predefined sentiment associated with it that is often good or bad or neutral. These lexicons are actually the core of the analysis. It is here that the general tendency of a text is determined depending upon the presence of certain words. For instance, words like "happy," "love," or "excellent" will fall into a category of positive words, and words like "sad," "hate," or "poor" will fall into a category of negative words. Some of the strengths of lexicon-based methods are that it is simple and can be deployed without using huge labeled datasets used for training, which makes them best suited to domain-independent sentiment analysis (Govindarajan, 2022).

  These methods are easy to use and generally involve a baseline technique in sentiment analysis applications. They come in handy particularly where one does not have access to domain-specific sentiment datasets or where it is impossible to develop machine learning models owing to time or computational constraints (Appel et al., 2016). However, despite their benefits, lexicon-based methods are very limiting. The most important disadvantage is that they are incapable of dealing with negation, a very complex linguistic feature. For example, a sentence like "I do not like this product" may be positive due to the word "like," but the negation part "not" makes it a negative overall sentence (Chavan et al., 2018).

  Lexicon-based methods typically have an extra problem: they do not appear to easily pick up on sarcasm and irony; that is, instances in which the literal meaning of the words does not convey the intended feeling. For example, the sarcastic phrase "Oh great, another delay!" probably gets coded as positive simply because of the word "great", but clearly expresses a negative sentiment (Govindarajan, 2022). These methods also suffer the problems of handling context specific meaning, where the sentiment associated with a word change with its usage. Since lexicon-based models do not understand the larger text context, they may also give out erroneous results especially when dealing with more subtle or complex datasets (Chavan et al., 2018).

Although lexicon-based approaches still remain widely used for many applications of sentiment analysis where speed and domain independence are the key, they are increasingly supplemented and even replaced by more sophisticated machine learning and deep learning models that are better equipped to deal with some of the complexity and richness of human language (Appel et al., 2016).

- **Machine learning-based approaches:** Machine Learning-Based Approaches: sentiment analysis relies on supervised learning algorithms to classify textual data. These are very sophisticated methods of lexicon-based methods as they rely on models that have been trained over labeled datasets. Algorithms found in this category are Naive Bayes, SVM and Random Forests. Each of the algorithms operates based on learning from massive amounts of labeled text data and then extracting relevant features to be applied in the classification of new unseen text.

  For instance, Naive Bayes uses probabilities to predict the category of the text as the chances that particular words are more likely to appear in that specific text (Agarwal et al., 2020). It is often known as simple and efficient compared with many other algorithms, particularly with large data sets. SVMs are particularly powerful especially in high-dimensional spaces like those of text classification where the data is represented as large feature sets, such as word frequencies or word embeddings. It is very sound at distinguishing clear lines between classes and is hence one of the most effective tools for sentiment classification in complex datasets (Agarwal et al., 2020).

  The ensemble of decision trees is used for making a prediction with Random Forests, offering another robust solution that avails against overfitting; instead, it retrieves much better results through aggregation of multiple trees' predictions, which gives even more stable outcomes (Govindarajan, 2022). Machine learning-based approaches generally outperform lexicon-based methods provided sufficient labeled data is available, but they're not without limitations. Generalization is one of the major problems: the models which have been working well on one dataset may fail when deployed to another domain, for example, if a model that has learned well about products performs poorly when generalized to political opinions. The reason for this is domain-specific language used, expression variations, and disparate distribution of sentiments in different datasets (Chavan et al., 2018).

  Thus, although machine learning approaches are more accurate especially when there is enough labeled data available, they do require careful tuning and other techniques like domain adaptation to make them generalize well to other contexts and domains (Govindarajan, 2022). Despite those challenges, they remain the preferred choice for the sentiment analysis in complex and high-dimensional datasets.

- **Hybrid approaches:** Hybrid approaches to sentiment analysis aim at combining the strengths of both lexicon-based approaches and machine learning methods for a more accurate and refined model of sentiment classification. These hybrid approaches draw on the initial scores learned from the lexicon-based approaches and further refine those with algorithms used by the machine learning models, thus closing the gap between the simplicity and efficacy of lexicon-based approaches and machine learning models. Lexicon-based methods are useful in quickly determining the sentiment of a piece using existing dictionaries, whereas machine learning algorithms make it even more complex: allowing the model to learn from data and accommodate more complex patterns of sentiment in text (Appel et al., 2016).

  For example, a hybrid model might start with an application of a sentiment lexicon on any given text to produce a crude polarity score, followed by passing this score to a classification procedure of some machine learning classifier like Naive Bayes (NB) or Support Vector Machines (SVM), fine-tuning the score based on context, negation handling, and domain-specific sentiment variations. Such a blend improves the baseline accuracy of sentiment detection but also enhances the model's ability to handle even more intricate linguistic features like sarcasm or context-based meaning (Chavan et al., 2018).

  Furthermore, sophisticated hybrid models often contain fuzzy logic, representing subtler intensities instead of categorically or in a binary manner. Fuzzy logic enables the model to operate with linguistic ambiguity and vagueness. In cases where sentiments are neither purely positive nor purely negative, but rather relatively ambiguous, fuzzy logic might be able to relate the intensity degree in a more subtle manner. Several studies have shown that hybrid models, especially those fusing fuzzy logic with sentiment lexicons, clearly have a high performance level beyond the performance of individual classifiers, such as NB and Maximum Entropy (ME) models. Hybrid approaches, especially when combining fuzzy logic with sentiment lexicons, are very effective in identifying otherwise silent sentiment polarities missed by traditional models (Appel et al., 2016).

Hybrid methods have been applicable to a range of tasks, from review analysis to social media content processing. The capacity to apply hybrid approaches to diverse contexts and datasets will certainly only enhance generalization as opposed to models that are only lexicon-based or only based on machine learning. Also, combining lexicon scores with machine learning refining provides a more resilient approach toward sentiment classification, and thus hybrid methods are pretty popular in modern research on sentiment analysis (Govindarajan, 2022).

- **Deep learning-based approaches:** The deep learning-based approaches in sentiment analysis have revolutionized the frontier of research as it makes possible the incorporation of the most complex architectures of neural networks: CNNs and RNNs in this kind of processing model. These models can capture syntactic, that is, structure information and semantic, that is, meaning information from textual data to make more nuanced and context-aware sentiment classification. Originally meant to be applied for image processing, the CNN is able to be sent into text data by considering the words within a sentence as pixels of an image, only capturing a sense of local patterns, such as phrases and n-grams that will eventually comprise the greater sentiment contained within a sentence. This is exactly why CNNs are so effective in producing aspects of a sentence that are purported to carry the sentiment such as positive or negative phrases without mastering the ability to fully understand the structure of the sentence itself.
RNNs are also well-suited for text analysis since it is designed to process sequences of data. That means the order of words plays a very important role in determining meaning, unlike image classification that doesn't depend so much on the order of the pixels. Variants of RNNs, namely LSTM networks, have been particularly successful in sentiment analysis as they can maintain memory of long-term dependencies and capture how earlier parts of a sentence influence the sentiment of later parts (Poria et al., 2020). For instance, in a sentence such as "I thought the movie would be terrible, but it actually turned out to be amazing," an LSTM can track the flip in sentiment from negative to positive. It is this ability to process long-range dependencies in text that makes RNNs and LSTMs invaluable for complex sentiment tasks.
These deep learning models have significantly improved the accuracy in sentiment analysis, mainly in noisy, complex datasets like social media posts, customer reviews, or tweets where traditional models often break down. Social media introduces some unique challenges, such as informal language, slang, and abbreviations that are challenging for models to handle other than the traditional word lists learned through deep learning models by learning contextual word representations. These models have an excellent ability to cope with variations in language most likely to confuse the lexicon-based or a machine learning-based model.
Moreover, deep learning techniques have been proved quite useful for finer-grained aspect-based sentiment analysis where sentiments related to specific aspects or attributes of a product or service are extracted (Poria et al., 2020). For instance, from an opinion about a hotel, an aspect-based sentiment model with the help of deep learning may find that the aspect "location" is positive while the aspect "customer service" is negative. This level of detail is crucial for any business entity to understand customer feedback beyond the general sentiment score.
The strengths of deep learning models go beyond just the ability to capture complex patterns in data. In this paper, their applicability to various domains and languages has also become significant, particularly when such deep models are brought together with pre-trained language models like BERT or GPT. Such models also carry the additional benefits of contextual embeddings, which capture meaning in words; such meanings are contextually derived depending on the surrounding text, for instance in subtle or domain-specific contexts, is invaluable for sentiment analysis (Poria et al., 2020).
In conclusion, deep learning-based approaches like CNNs and RNNs have currently advanced as far as improving accuracy, handling complex data, and offering deeper insights into specific aspects of sentiment in making them truly indispensable tools for modern sentiment analysis tasks.
- **Fuzzy logic and rule-based approaches**: One of the state-of-the-art techniques to deal with the fuzziness and vagueness inherent in human language is the Fuzzy Logic and Rule-Based Approaches, handling sentiments that were otherwise classified as static categories, either positive or negative, or at best even neutrally balanced. Instead, fuzzy logic assigns degrees to sentiment, which are known as fuzzy membership degrees. This approach allows for a more accurate interpretation of the intensity of sentiments, which may either be very positive or mildly positive or from neutral to slightly negative. Fuzzy logic can capture the subtlety of human emotions and opinions without imposing strict binary categories for these sentiments (Jefferson and Roberts, 2017).

Such approaches are very useful when dealing with semantic ambiguity, that is, with regard to the context, the sentiment of a word or expression might or might not be the same. For example, "fine" could be positive in one perspective yet lesser on another as compared to in sentences like "The service was fine" yet somehow less enthusiastic when saying "The product was just fine". Fuzzy logic makes it versatile and capable of interpretation as it uses degrees rather than treating all instances of the word "fine" as strictly positive or neutral.

Another refining of fuzzy logic-based models includes the rule-based systems; they increase yet another level in refinement sent into estimation. Rule-based systems allow the model to continually update its sentiment classification based on predefined rules incorporating context or specific conditions. For instance, it may be: negation ("not happy"), intensification ("very happy"), and the score for the sentiment would shift accordingly. It merges rule-based logic with fuzzy sentiment analysis into a formidably powerful technique that would be capable of controlling complex linguistic constructs, such as sarcasm or idiomatic expressions, into which simple modeling techniques get stuck (Jefferson and Roberts, 2017).

Fuzzy logic is also useful in domain-specific sentiment analysis where some words have varying connotations. For example, the word "volatile" can have a neutral or even positive connotation depending on the context used in the financial text. A fuzzy logic system can assign degrees of sentiment based on these nuances, and hence the analysis will result in more accuracy and finer granularity than those binary or categorical models (Chavan et al., 2018).

Overall, fuzzy logic-based models enhance the granularity of sentiment analysis because varying degrees of sentiment intensity are allowed, and results are also tuned with context-specific rules. Hence, they should be best applied when detailed sentiment understanding is required, such as in product reviews and consumer feedback or even social media analysis, where language tends to be often less formal and more context-dependent.

6. SENTIMENT ANALYSIS IN REAL-WORLD SCENARIOS

Sentiment analysis helps to process and interpret voluminous amounts of unstructured text, and it does have very crucial applications in various spheres by providing critical insights into customers' opinions, emotions, and attitudes.

- **Marketing and e-commerce**: Sentiment analysis is one of the most widely used applications of marketing and electronic commerce domains. Businesses here use sentiment analysis to grasp customer sentiment about products and services so as to facilitate maximum satisfaction in customers and optimize their product development. This can be performed on Amazon or other social networking sites where sentiment analysis essentially helps companies evaluate how the customers feel about their offerings through product reviews or other comments posted online (Wu et al., 2017).
  Domain-specific sentiment classifiers trained and sensitized according to unique types of expressions which elicit product-specific sentiments can give companies a personalized marketing strategy and promote better relationships with their customers. For instance, just through the recurrence of repetitive positive or negative sentiments from reviews, businesses would alter their message or change features of a product to live up to expectations. Sentiment analysis also allows a company to monitor real-time branding aspects by determining the sentiments articulated on social media, hence enabling companies to react to customer feedback on time and interact better with their consumers.
  Sentiment analysis in the online shopping scenario helps immensely to identify customer behaviour and solve major pain points. Personalized product suggestion based on sentiment-emitted data from previous customer interaction is one such application where such type of analysis can be employed. E-commerce websites can calibrate their suggestion algorithms by observing reviews and ratings. This also leads to more personalized and fulfilling user experiences, which in turn can generate more sales and retainment (Quan and Ren, 2016).
- **Finance and stock market prediction:** The most important tool identified by financial institutions and investors who may understand market sentiment and predict changes in stock price is sentiment analysis within the finance sector. Analysing huge volumes of textual data from sources like news articles for which financial news is published, posts from social media, and economic reports yields very valuable insights into how the public perceives economic trends, stock performance, and general investor sentiment. This

information is very imperative to the financial analyst tracking the trend of change in the tone of the sentiment to predict possible market fluctuations in the future (Dong and Liu, 2021).

Researchers have developed industry-specific sentiment lexicons for the financial sector, so analysts can further calibrate their sentiment models to capture financial jargon and context-specific meanings. This kind of domain adaptation would strengthen the ability to monitor shifts in market sentiment that might signal changes in stock prices or volatility in the economy. For instance, sharp rise in negative feelings toward a particular stock or any economic event may signal an impending decline whereas positive feeling could show a higher jump in the prices of stocks.

- Sentiment analysis not only helps in identifying trends in general mood in the market but also helps in identifying patterns of market optimism or pessimism. This is very crucial for a well-informed investment decision. Predictive models that include sentiment data are known to greatly enhance the precision of financial forecasts. It hence enables financial institutions to respond to fast shifts in the market and strategically gain an advantage (Consoli et al., 2020). This ability will provide real-time insight into the sentiment, which will help investors improve their assessment of the market condition and enable them to make necessary adjustments on the portfolio that may help in reducing risk and taking advantage of emerging opportunities.
- **Healthcare:** In the healthcare sector, sentiment analysis is an increasing tool used to gauge patient experiences, mainly through online reviews and feedback about healthcare providers and institutions. It can point out main issues-through textual data from a site where patients share experiences-by processing their data, whether care quality, staff behaviour, or facility conditions. This information provides valuable input health care practitioners with regard to patient satisfaction and enables them to rectify specific issues to increase their overall standards of care. For example, if the common grievance regarding waiting time or overstaffed interaction typified the unwarranted dissatisfaction, healthcare institutions can make necessary changes for better patient experience.

  In addition, social media platforms like Twitter and Facebook are used in analysing public mood concerning vital health issues like diseases, vaccines, and health policies. This is real-time direct sentiment analysis and health authorities can gauge the public mood on the same and, therefore, change their communication strategies. For instance, during the pandemic, it can help identify misinformation, understand the reasons for vaccine hesitancy, and even tune campaigns to receive more trust and compliance concerning public health recommendations (Islam and Zibran, 2018).

  One critical application of sentiment analysis in healthcare concerns the domain of mental health. By analyzing the conversations of patients either in clinical settings or online forums, sentiment analysis can discover early signs of mental health distress, such as depression, anxiety, or suicidal thoughts. Such information could be discovered by a trained machine learning model on health care applications that can identify emotions or patterns of distress requiring immediate clinical intervention. The proactive approach adopted in the case can effectively allow healthcare professionals to assist individuals at risk before things get worse. In this context, sentiment analysis will become one of the emerging frontiers in the adoption of AI for the betterment of patient care and mental health monitoring.
- **Politics and public opinion:** Sentiment analysis has become an indispensable tool in politics and public opinion for political campaigns and governmental organizations searching for what the public thinks about a particular policy, who should be voted into office during the election period, and what the populace is arguing over on political matters. It does so through the analysis of textual data generated on social media platforms, news articles, and blogs. This data proves particularly useful in predicting election outcomes through monitoring trends in sentiment over time. For example, a spike in positive sentiment towards a certain candidate or policy could sign onto an upward trend in public support, whereas negative could often be taken as a sign of potential strategy modifications.

  Sentiment analysis goes beyond just predicting elections. It tracks political rhetoric on issues that have either been debated or are sensitive, giving the government a data-driven understanding to make decisions more in tune with public position. It gives insight into how people feel about new laws, economic measures, and international relations to allow policymakers to know how effective their policies are and change their communication approach to relate better to people. Through the analysis of these feelings, therefore, governments can also enact policies that benefit their citizenry's likes and dislikes.

Even further, sentiment analysis could be a powerful tool against propaganda and the spread of misinformation-a phenomenon that has really taken on the social media platform these days. Since patterns of false and misleading information can be established, sentiment analysis helps governments and organizations to respond more prudently towards disinformation campaigns that can distort public opinion or make voters behave in a particular way. The capability maintains the integrity of political processes so that public opinion is not swayed by wrong or false information (Lyu et al., 2023).

- **Education:** Sentiment analysis is increasingly applied in the education sector: from garnering students' feedback over courses or experiences at the university to discussions on social media about higher learning institutions. Now, universities and online educational platforms are increasingly using sentiment analysis to follow up on how students feel about their experiences, which areas they need to improve on, and how to better learn. Institutions can make use of student evaluations or reviews through feedback questionnaires to analyse feelings-both positive and negative. This helps learning institutions to better understand effective and less effective components of the learning process. This knowledge also assists schools and universities in tailoring their services, curricula, and systems for support to adequately address the needs of students in improving their learning outcomes (Zhou and Ye, 2020).

  Similarly, with the aid of sentiment analysis, institutions can discover bigger trends in student satisfaction over time, allowing them to inform changes in programming, course development, or campus services. For instance, a general negative comment shown through sentiment analysis regarding the availability of certain support services indicates that the university should focus their improvements in this space for an enhancement of the student experience.

  Another is the integration of sentiment analysis, which measures the emotional student response to online learning material or assignments. These technologies allow the educators to analyse students' current response in real time, hence providing timely critical input concerning the engagement level. If the sentiment analysis reveals that students fail to grasp certain topics or assignments, instructors can support them or access related learning resources early enough. This proactive approach ensures students don't get behind and that all support is provided before falling behind. Additionally, sentiment tracking in online discussions or assignments can also identify areas where students are highly engaged, so repeated successful teaching strategies can be applied elsewhere in the curriculum.

  Generally, sentiment analysis in education allows for an improvement in a learning environment, given a better understanding regarding the emotional and academic experiences that students undertake.

- **Software engineering:** In software engineering, sentiment analysis had become an important tool in trying to understand what developers were saying by analysing reviews of programming languages, interpreting sentiments expressed in bug reports and technical documentation, among others. Given the unique linguistic problems technical language poses for software developers, domain-specific tools for sentiment analysis have consequently been developed. All of these tools are designed to be familiar with the specific vocabulary, jargon, and tone that developers use in communication; therefore, the analysis of these sentiments becomes even more precise.

  For example, analysing bug reports may help evaluate the degree of irritation or satisfaction of developers with specific tools or features, providing information that closed-ended feedback tools are not likely to obtain. Sentiment analysis enables companies to measure the effectiveness of their software development practices and monitor the most common pain sources of the latter. Most attractive applications of these techniques are open-source projects, as developer sentiment itself - in the form of code reviews, bug reports, and community discussions - may represent an extremely efficient 'sensor' system for important issues, giving maintainers a chance to know about problems as soon as possible and react to them immediately (Islam and Zibran, 2018).

  Sentiment analysis further is applied to monitor the happiness of developers with a number of tools, libraries, and frameworks that is an essential task for the companies developing the software to improve on their own. From this point, through the analysis of sentiments regarding technologies on developer communities, a company can decide what software feature needs refinement, or where a new software feature can be offered as a filler for a missing feature which developers expect. For instance, systemic negative remarks about specific aspects of a particular library may cause companies to begin releasing patches or even updates to try and solve the problems behind those issues. Meanwhile, users' responses about specific features may be able to influence the future course of development in a way that best suits their needs.

In many ways, companies can come to understand how the software is received, and in turn can improve on it based on feedback from users by tracking sentiment that crosses developer forums, GitHub repositories, and technical documentation. Through sentiment analysis, in a way, valuable insights are provided to enable quicker, better, and more informed decisions when developing software with regard to better quality and helpful user experiences in the long run.

7. DATASETS USED IN SENTIMENT ANALYSIS

Availability of many datasets that consist of text in diverse languages, domains, and platforms has made a massive difference in the progress of research done in sentiment analysis. These datasets become basic building blocks for creating and testing sentiment analysis models, most importantly in supervised learning-based tasks. These datasets contain the following types:

- **IMDb dataset:** IMDb dataset contains a lot of movie review data and thus is one of the most widely used datasets for sentiment analysis tasks. This dataset has labeled data where reviews are classified as either positive or negative, best suited for training and testing most machine learning models like Naive Bayes, Logistic Regression, and Random Forest. These models can be experimented over this dataset to predict the sentiment polarity, thus giving a benchmark for comparison of performance by different models (Tripathi et al., 2020). Apart from this, it is often used in combination with techniques like word embedding, GloVe (Global Vectors for Word Representation), as it converts textual data into numerical vectors, through which models understand semantic relationships between words. These embeddings improve the performance of the models for sentiment analysis, in consideration of a review's context while increasing accuracy. Using word embeddings and also labelling on IMDb has placed it as the corner stone for most the work done in researching sentiment analysis when evaluating different approaches to machine learning and deep learning architectures.
- **LABR dataset:** LABR is a very large Arabic book review dataset for sentiment analysis. It is the largest available dataset so far with 63,000 book reviews for Arabic sentiment analysis. LABR supports two major tasks: sentiment polarity classification under binary labels and ratings classification on a 1-5 star scale. This dual utility makes LABR a highly versatile dataset, whereby researchers can run relatively simple sentiment classification tasks and very fine-grained classifications related to rating prediction tasks (Nabil et al., 2014). Due to the richness of diversity in the LABR dataset, it has emerged as a benchmark for Arabic sentiment analysis studies. Given the relatively smaller number of large datasets available in the Arab world as compared with other languages, such as English, LABR represents a crucial asset which may further push the research into the world of sentiment analysis in the Arabic language. Its gigantic size and diverseness of opinions provide possibilities to develop and fine-tune machine learning models that are tuned to the complexities of the Arabic language, such as richness in morphology and varieties of dialects. Researchers use LABR who rely on various techniques, such as word embeddings, feature extraction, machine learning algorithms, to explore how to classify sentiments in Arabic texts. The dataset considerably enhanced the accuracy of Arabic sentiment analysis systems and is widely cited in the process of developing models for natural language processing in Arabic. The dataset is useful when testing the models' robustness against linguistic style and in variations in the intensity of sent items. This could, for example, be understanding the correctness or otherwise of these models when generalized to real, unstructured textual data.
- **The Turkish twitter sentiment dataset:** The Turkish Twitter Sentiment Dataset is a valuable resource for sentiment analysis in Turkish: manually annotated Turkish tweets. As one of the first ever datasets specifically dedicated to Turkish, it addresses some of the unique challenges that social media text poses-most notably slang, abbreviations, and misspellings. Social media tools like Twitter are also known for shortforms and slang usage because they have character limits. It will allow researchers to construct better sentiment analysis models capable of effectively parsing such non-standard forms by using Turkish tweets (Köksal and Özgür 2021). This dataset would be very suitable for modern transformer-based approaches, such as BERTurk that is a pre-trained BERT model for Turkish. It has been used to test and tune a wide range of models, including BERTurk, hence this is quite relevant for transformer-based sentiment analysis. Such models with deep learning architecture are well-suited for dealing with relationships between words within context, thus doing justice to the fragmented and informal language often present in social media. The Turkish Twitter Sentiment Dataset therefore offers hand-annotated data for the enhancement of the

performance of tasks relating to Turkish language NLP, including sentiment analysis itself, and bridges the gap in resources available in regard to research into Turkish language AI. Today, researchers count on this dataset as a source of developing more robust models that can further manage the complexities of Turkish social media text, while continuing the task of adding to the competitiveness of research into the sentiment analysis of underrepresented languages.

- **Arabic Sentiment Tweeta Dataset (ASTD):** This dataset is among the most important in the area of Arabic sentiment analysis, particularly on text coming from social media. It encompasses about 10,000 manually annotated Arabic tweets in four categories: objective, subjective positive, subjective negative, and subjective mixed. In this way, fine-grained analysis of sentiment becomes feasible, enabled by distinguishing between states of fact and opinions, as well as positive, negative, or mixed sentiments. The granularity makes ASTD particularly suitable for tasks greater than simple binary sentiment classification (Nabil et al., 2015). ASTD is widely used as a benchmark dataset for sentiment analysis in Arabic social media contexts. Social media platforms are a rich source of user opinions and reactions on social issues, and thus the sentiment analysis of this data is considered crucial for the understanding of public opinion in the Arab world. analysing Arabic tweets is particularly challenging in terms of richness in morphology, dialectal varieties, and the informal characteristic of social media language. Annotated data from ASTD bridge this gap as they support the development and testing of machine learning-based models adapted for Arabic sentiment analysis. Researchers have tried many different machine learning algorithms with ASTD testing: Naive Bayes, SVMs, and even more complex deep learning models. In terms of providing benchmark results, ASTD was the standard reference point for testing and comparing the performance of different techniques and architectures when it comes to sentiment analysis. It has become part and parcel of the dataset while advancing research in Arabic sentiment analysis, particularly in its social media applications.
- **SentNoB dataset:** We focus on SentNoB-dataset: it is a highly useful resource for the task of Bangla sentiment analysis. The data is composed of informal public comments through news articles and videos-it encompasses all different possible opinions and sentiments with variety in everyday language-the 13 different domains are covered by SentNoB. It includes various Bangla dialects and informal speech and is also somewhat of a noisy data-set. Such noisy data arising from dialectal variations, colloquialism, and grammatical errors present a challenge for the sentiment analysis models, since these are common phenomena in user-generated contents (Islam and Zibran, 2018). The emphasis of managing such noise data, thus reflecting real-world language use, perhaps differentiates SentNoB, since Bangla happens to be a morphologically rich language with multiple regional dialects in its usage. The dataset has been used in experimenting along with different approaches for sentiment classification. What is more interesting in this dataset, however, is that the hand-crafted lexical features-outperformed neural networks in this dataset with manually designed rules and dictionaries. This is remarkable, as it indicates that for some types of data, state-of-the-art deep learning architectures are inferior to traditional methods, especially for noisy and informal data. Of course, advancing research on sentiment analysis for the Bangla language, with its limited resources, is now possible with datasets like SentNoB. It lays a foundational platform with which such machine learning models could be built and tested that may support formal as well as informal Bangla texts, and hence it presents itself as an important tool for researchers and practitioners to be used in the concerned domain.
- **NoReC_fine dataset:** NoReC_fine is a rich source dataset directed for fine-grained Norwegian sentiment analysis, very comprehensive in the domains addressed which include literature to movies and various products, a fully versatile dataset for the cross-domain application of sentiment analysis in every sector. A feature of this dataset, NoReC_fine, that sets it apart from any other dataset is the level of detail with which the labels are annotated, detailed identification of polar expressions, targets, and opinion holders. This resolution is very useful for aspect-based sentiment analysis, where in addition to positive or negative classification, the sentiment can be related to specific features or aspects of a product or service (Ovrelid et al., 2019). For example, on a product review, researchers can extract what aspect is being referred to-whether it is "battery life," "camera quality," or anything else-and with what sentiment that aspect is being regarded. This feature makes NoReC_fine a very valuable resource for business users and researchers eager to find more in-depth information about customer opinions at a much greater level of depth. Additionally, the structure of the dataset includes annotations of opinion holders that would permit analysis of who is speaking and in what context. NoReC_fine was heavily put to use in performing the performance evaluation of aspect-based sentiment analysis models and helped to contribute some amount of innovation in terms of

improvement in the field of sentiment analysis for less-resourced languages like Norwegian. Its fine-grained approach makes it excellent in building models that need to go beyond simple polarity classification and into specifics like what exactly people like or dislike about a particular entity.

8. EVALUATION METRICS

The evaluation of models in sentiment analysis is critical to determine their effectiveness in the classification and interpretation of sentiment. Different metrics and benchmarks are used to judge the performance of these models. What follows is a brief overview of some of the key evaluation metrics and benchmarks used for the task of sentiment analysis:

- **Accuracy:** Accuracy is the most used metric to measure the performance of sentiment classification tasks. It just refers to the number of correct predictions over all. There is no problem with the calculation and interpretation of accuracy metrics, although accuracy may not always be reliable when dealing with imbalanced datasets where, for example, one sentiment class might dominate the data. Then, accuracy can overestimate the performance of the model in these cases since it will favor the dominant class (Ribeiro et al., 2015). Therefore, for further balanced evaluations, usually, other metrics like precision and recall are considered.
- **Precision, Recall, and F1 Score:** Precision refers to how many true positives there are amongst all the positive predictions that have been made. Recall, or sensitivity, involves how many true positives are identified out of all actual positive instances, focusing on what the model can capture of all relevant sentiment instances. F1-score is one kind of harmonic mean of precision and recall, hence it provides a well-balanced measure regarding the performance of the model. However, if the dataset is imbalanced, then the F1 score really proves helpful since it considers both the false positives and the false negatives (Nazar and Bhattasali, 2021).
- **Area Under the Receiver Operating Characteristic Curve (AUC-ROC):** Another assessment for measurement that is tested within the scope of classification problems, particularly for a classification problem which is binary, is known as the Area Under Curve-Receiver Operating Characteristic, or AUC-ROC. In this case, the curve of ROC plots the true positive rate against the false positive rate at different threshold settings; the AUC represents the area under this curve. AUC-ROC gives a better elaboration about how the model works by considering its ability to separate the classes at different thresholds for imbalanced datasets (Alaei et al., 2019).
- **Kappa Statistic (Cohen's Kappa):** Cohen's Kappa estimates the correlation between the classifications of predicted emotions and the actual labels after adjusting to the possibility for a chance agreement. This is an important measure of comparison for multiple classifiers or inter-rater agreement in tasks such as aspect-based sentiment analysis, as different people annotating the same sentiment might have interpreted it differently. Cohen's Kappa goes beyond simple accuracy which only gives a perception of what it actually performs (Potts et al., 2020).
- **Cross-Validation (K-Fold):** K-Fold Cross-Validation is an effective method to test the model performance based on a dataset that can be divided into K subsets, or "folds." The model is trained K-1 times, while the remaining fold is used for testing. This is carried out K times in order to have a robust evaluation against different partitions of the data. Overall, the risk of overfitting is reduced in the technique because it tests the model performance against different subsets of data, which yields a more accurate and generalizable estimate of its performance (John and Kartheeban, 2019).
- **Entropy-Based Metrics:** Entropy-based metrics have also recently been proposed for model selection in sentiment analysis. These measure uncertainty in the model's predictions and thereby shed light on how the consistency and reliability of the model's outputs increase with its level of confidence. Indeed, what entropy-based metrics complement well with traditional metrics like accuracy are more precise indications of how confident a model is in its predictions, especially if the model's decisions are to be applied to critical areas (Valverde-Albacete et al., 2013).

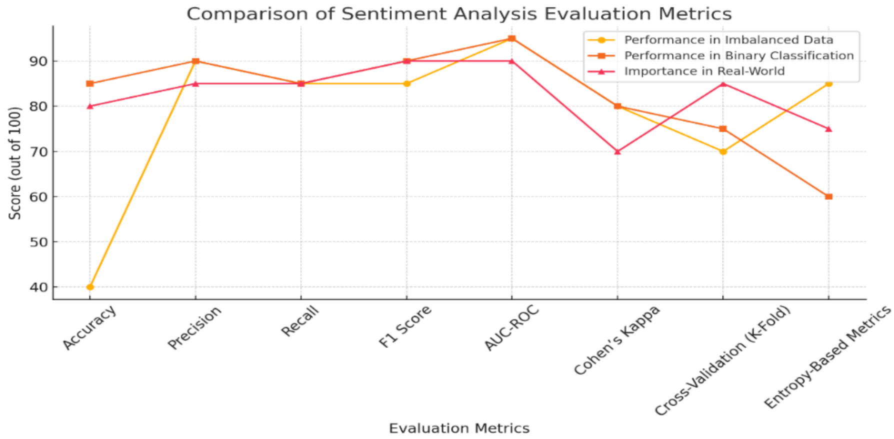

Fig 1: Comparison of Sentiment Analysis Evaluation Metrics

In order to elaborate on how different evaluation metrics may be visualized for good performance across key dimensions like imbalanced data, binary classification, and real applications, the performance of such evaluation metrics is demonstrated in Fig 1. Elaboration given below:

- **Performance in imbalanced data:** Other measures such as Precision, Recall, F1 Score, and AUC-ROC function well when they have to deal with imbalanced datasets, i.e., one class considerably dominating the other. These are useful measures since they take care of the ability of correct classification for the minor classes. However, the accuracy measure is fairly misleading in such scenarios since the system will always overestimate since it favours the majority class.
- **Performance in binary classification:** From the list of metrics that would be useful for this type of problem, binary classification tasks distinguishing between positive and negative sentiments, AUC-ROC, Precision, and F1 Score were the most descriptive of the ability of the model to distinguish between the two classes of sentiment, far more subtle than accuracy because it merely computes the percentage of accurate predictions without considering the subtlety of the decision making in the context of binary classification.
- **Importance in real-world applications:** Measures that are much more appreciated in sentiment analysis real-world applications - such as for customer feedback or predicting the stock market - include Precision, F1 Score, and AUC-ROC. These metrics give a much better understanding of the performance of the model beyond accuracy with counts of false positives and false negatives, which are very important in real decision-making processes.

Fig. 1 reveals that, although accuracy is the most commonly used metric for evaluation because of its interpretability and ease of use, more advanced metrics are Precision, Recall, F1 Score, and AUC-ROC, which altogether give a more dramatic view of model robustness especially when faced with imbalanced datasets as well as when performing binary classification tasks. This enables the researches and the practitioners to have more in-depth insights about the performance of the models, leading to more complete assessments than when using several different datasets and contexts.

These metrics can be used to harden the evaluation of the models that perform sentiment analysis; they provide a baseline for inspecting accuracy-related limitations and real-world data hurdles.

Table 1: Comparing Sentiment Analysis Approaches on the basis of Evaluation Metrics:

Approach	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Cohen's Kappa	Reference
Lexicon-Based	Low	Moderate	Low	Low	Low	Low	(Govindarajan, 2022)
Machine Learning-Based	High	High	High	High	High	High	(Agarwal et al., 2020)

Hybrid Approaches	Higher	High	High	High	High	Higher	(Appel et al., 2016)
Approach	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**AUC-ROC**	**Cohen's Kappa**	**Reference**
Deep Learning-Based	Highest	Highest	Highest	Highest	Highest	Highest	(Do et. al, 2019)
Fuzzy Logic & Rule-Based	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	(Jefferson and Roberts, 2017)

Table 1 provides a comparative summary of the performance of different sentiment analysis methods across key evaluation metrics, highlighting their strengths and weaknesses in terms of accuracy, precision, recall, F1 score, AUC-ROC, and Cohen's Kappa. Deep learning approaches are shown to consistently outperform other methods across all metrics, making them the most powerful and accurate for sentiment analysis tasks, especially for large and complex datasets. These methods excel in capturing context, handling sarcasm, and dealing with imbalanced data, making them ideal for advanced sentiment analysis applications. The lexicon-based methods are simpler to implement and, respectively, have worse performance across the metrics. It is because they do not manage the sophisticated linguistic features such as sarcasm, meaning from context, and negation well. At the same time, simplicity makes lexicon-based methods useful for very fast and domain-independent sentiment analysis. Machine learning-based methods and hybrid approaches demonstrate more performance than that of lexicon-based methods, mainly if they are trained on a large labeled dataset. Hybrid approaches that combine lexicon-based approaches with machine learning algorithms perform well on all three metrics as they utilize the strengths of both approaches. Fuzzy logic & rule-based approaches are middle ground and useful for the task of handling ambiguity in sentiment; however, they tend to be underwhelming in performance comparing to deep learning-based methods, particularly in large-scale applications.

Overall, from table 1, it is vividly clear that deep learning approaches show better accuracy while other methods may suit the specific requirements and complexity of the sentiment analysis task.

9. RECENT CHALLENGES IN SENTIMENT ANALYSIS

Despite significant progress in this sphere, sentiment analysis still faces many challenges that determine its accuracy, generality, and broad applicability of the model across multiple domains. Some of the major ones include:

- **Handling sarcasm and irony:** Sarcasm and irony are huge challenges to sentiment analysis models as these modes of communication frequently use positive words to express negative sentiments or vice versa to eventually mislead simple sentiment polarity detection-based models. For example, "Oh great, another delay!" is a sarcastic statement with positive words like "great" but carries negative sentiment. Traditional machine learning models, which rely on the literal meaning of words for the most part, fail to capture the intended sentiment in such cases (Mohammad, S., 2017). Developing models that can detect sarcasm is still an open challenge in the field.
- **Sentiment quantification:** Quantification of sentiment is the estimation of how often such sentiment classes occur in large text corpora. Simple classification schemes tend to over- or underrepresent the overall sentiment of the data, since they're biased by higher values for false positives than false negatives. For example, when tens of reviews are misclassified as positive that actually are negative, the overall perceived sentiment of the data likely gets skewed. Models specifically tailored for quantification rather than simple classification are surprisingly under-represented in sentiment analysis research (Sebastiani, 2016).
- **Cross-domain generalization:** One of the most challenging issues in sentiment analysis is cross-domain generalization: models trained on one dataset that don't do very well in their application to a different domain. It results from the difficulty of language usage in different domains and variations in the distribution over sentiments. For example, a movie review trained sentiment model does not perform well when applied to product reviews in view of the difference in how the sentiments are expressed. Another area to which

research is still ongoing includes model development that generalizes across domains (Raghunathan and Saravanakumar, 2023).

- **Multimodal sentiment analysis:** It has been a long era for sentiment analysis in which the use of multiple content types of social media and online platforms has increased the chances to get more and more challenges of multimodal sentiment analysis. In other words, multimodal sentiment analysis is the integration of text, images, video, and other types of data which enhance accuracy in sentiment classification. Even if the integrations of these different types of data give more functionalities, it introduces complexity in the design and training phase of the model. Combining the textual, visual and auditory signals for the prediction of coherent sentiment is still an open problem (Kumaresan and Thangaraju 2023).
- **Handling dynamic and real-time data:** The rise of social media platforms like Twitter and Facebook presents the challenge of processing massive amounts of data in real-time. Dynamic and real-time sentiment analysis is particularly important during events like political elections, crises, or product launches, where opinions and sentiments shift rapidly. However, developing models capable of scaling and processing data streams in real-time remains a difficult task, as these models need to balance speed and accuracy. Scalability and efficient data processing are essential for successful real-time sentiment analysis (Ebrahimi et al., 2017).
- **Cross-lingual sentiment analysis:** Cross-lingual sentiment analysis involves extending sentiment analysis models to multiple languages, which presents challenges related to grammar, syntax, and cultural nuances. Most sentiment analysis models are developed primarily for English, and adapting them to other languages requires overcoming significant linguistic differences. Cross-lingual sentiment analysis models are crucial in today's globalized world, where sentiment data is generated in many languages, yet much research is still needed to address this challenge effectively (Pozzi et al., 2017).
- **Data imbalance and sparsity:** Imbalanced datasets, where one sentiment class (e.g., positive) dominates the others, can lead to biased models that overpredict the majority class. Similarly, data sparsity, particularly in certain languages or domains, limits the ability to train effective models. These issues can severely affect the performance of sentiment analysis systems. Techniques like oversampling, undersampling, and transfer learning are often employed to address data imbalance and sparsity, but further advances are required to overcome these challenges fully (Shayaa et al., 2018).

Addressing these challenges will be critical for improving the accuracy, scalability, and adaptability of sentiment analysis models, enabling their broader application across different languages, domains, and real-time contexts.

10. LEVERAGING MULTIMODAL SENTIMENT ANALYSIS TO ADDRESS COMMON CHALLENGES

Since its time of origin, sentiment analysis has faced numerous challenges that have made the process of emotion identification inaccurate and unreliable. The most traditional approaches to unimodal methods rely mainly on a single source of data, that is, either text-only, making their efforts in offering an effective description of human subtle emotions a failure. This dependency often leads to misunderstandings, especially concerning a phenomenon like sarcasm, fuzzy language, or the absence of non-verbal signs, which significantly impacts sentiment (Poria et al., 2017; Cambria et al., 2017).

Multimodal sentiment analysis has emerged as a stronger solution to such problems. With the incorporation of multiple data modalities that are comprised of text, audio, and visual cues, this approach offers a more holistic view of human expression (Zadeh et al., 2017). These modalities contribute different contextual information, respectively: linguistic content is presented in text, intonation and pitch in audio, and facial expressions and body language can be found in visual data. This, in turn, has strong enhancement for the robustness of sentiment interpretation, permitting systems to detect emotions with greater accuracy (Morency et al., 2011).

Multimodal analysis is going to make a better cover over the issue of sarcasm or ambiguous language which usually confuses unimodal systems. For example, although in black and white on paper a sarcastic remark can be positive, further cues like tone of voice or facial expressions might well expose it towards its actual meaning (Poria et al., 2019). This kind of deep analysis bridges the gap between being computational and close to humans in understanding: it is much closer to the way people intuitively perceive and interpret emotions.

Besides this, multimodal systems show robustness against noisy or partially missing data. When a single modality cannot give some useful information due to noise, say audio quality is low, another modality may have some relevant data available to complement this modality so that the whole analysis process becomes accurate (Baltrusaitis et al., 2018). This kind of multimodal integration will work in highly interactive applications, like the real-time monitoring of sentiment on multimedia content and empathetic responding AI applications that care for user's emotional states (Zadeh et al., 2018).

To further verify the performance of unimodal and multimodal techniques, we studied some of the recent work done in this field and compared their efficiency. Table 2 below summarizes the unimodal and multimodal techniques used in the analyzed references.

Table 2: Comparison of Unimodal and Multimodal Techniques

Reference	Unimodal Techniques	Multimodal Techniques	Fusion Method
Xu (2023)	Text-Based	Text + Audio + Visual	Early Fusion
Zhang et al. (2023)	Audio-Based	Text + Visual	Hybrid Fusion
Kaur and Kautish (2022)	Text-Based	Audio + Visual	Decision-Level Fusion
Chandrasekaran and Nguyen (2021)	Text-Based	Text + Audio	Early Fusion
Mahersia and Ezzameli (2023)	Audio-Based	Text + Visual	Late Fusion
Das and Singh (2023)	Visual-Based	Audio + Visual, Text + Audio + Visual	Hybrid Fusion with Attention
Gandhi et al. (2023)	Audio-Based	Text + Audio + Visual	Transformer-Based Hybrid Fusion
Sun et al. (2023)	Visual-Based	Text + Audio	Transformer-Based Dual Fusion

A comparative study on the efficiency of various unimodal and multimodal techniques discussed in the references in Table 2 is made. This comparison is presented in the form of three different charts as shown in Fig 2, Fig 3 and Fig 4. It is observed that maximum efficiency achieved with unimodal techniques is somewhere around 80% with text-based analysis as depicted in Fig 2 whereas, multimodal techniques showed a significant improvement in the maximum efficiency as shown in Fig 3 and Fig 4.

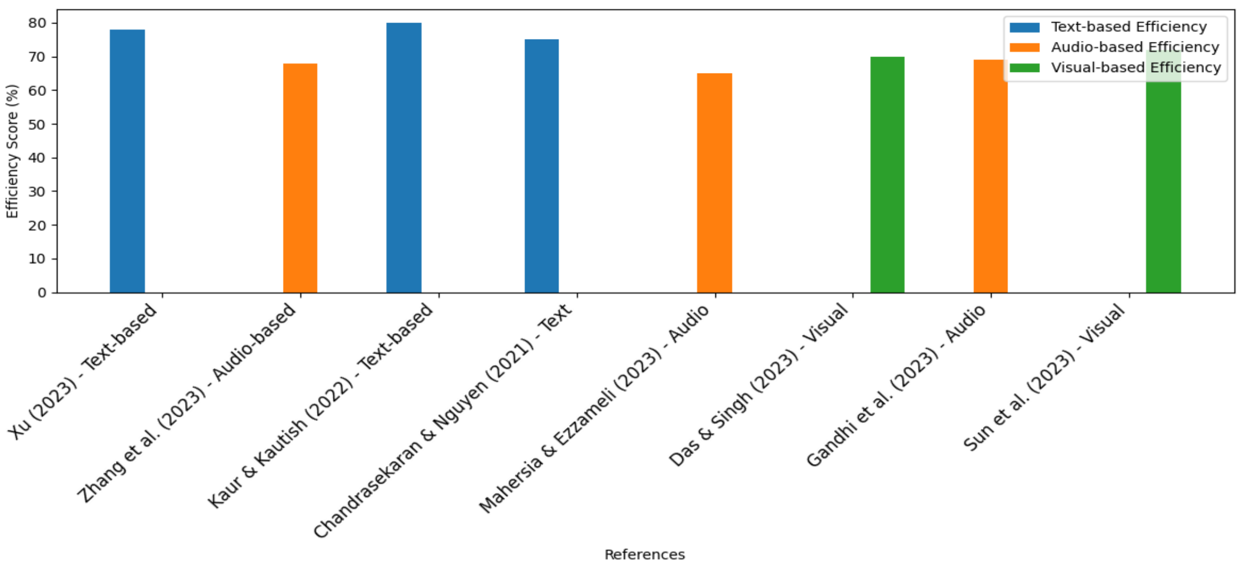

Fig 2: Comparison of unimodal sentiment analysis techniques

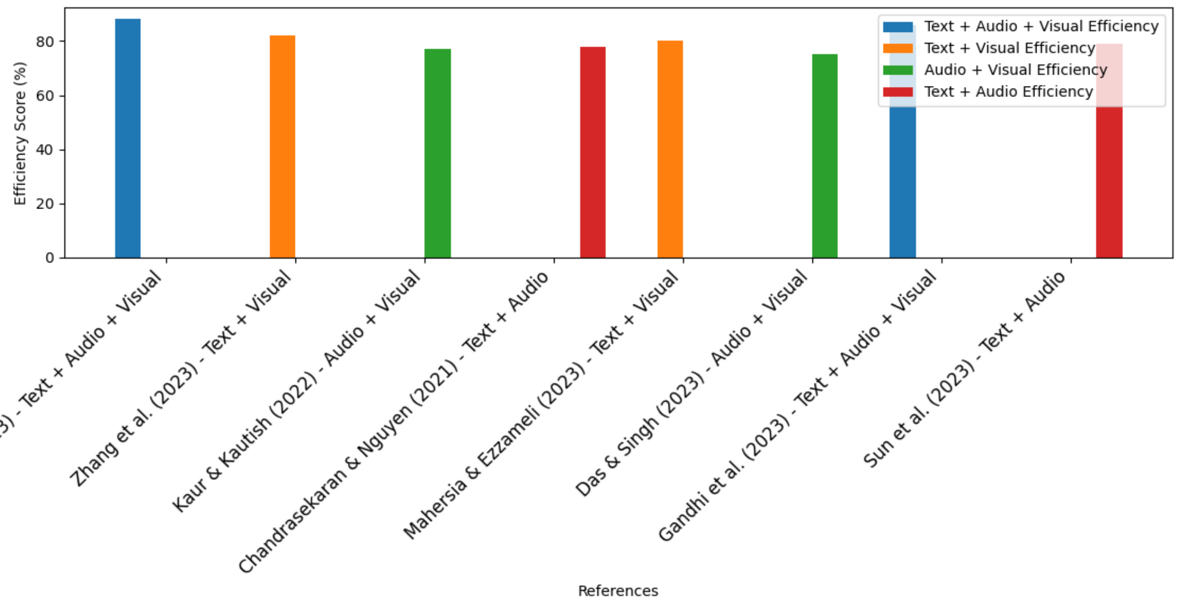

Fig 3: Comparison of multimodal sentiment analysis techniques

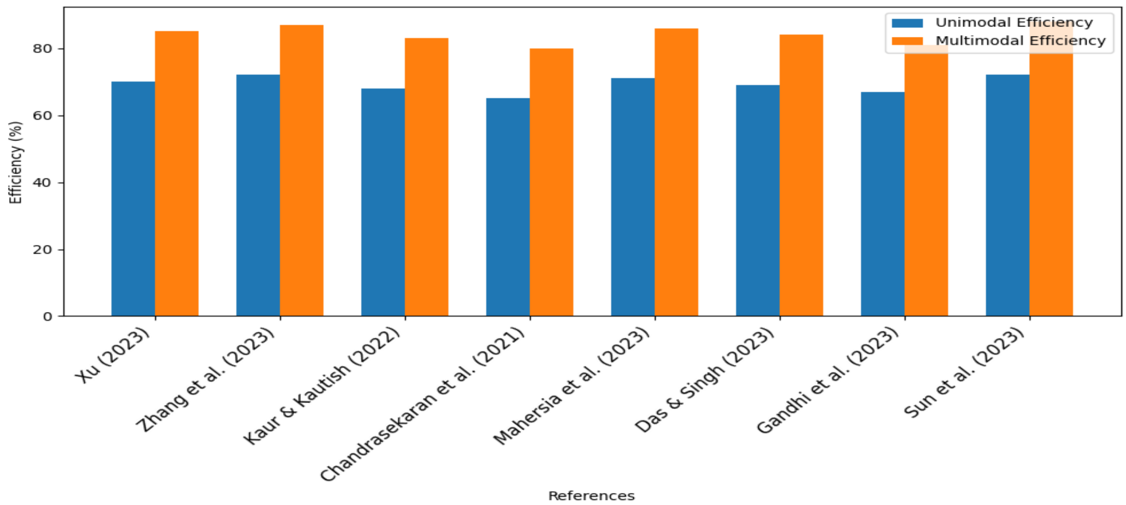

Fig 4: Efficiency Comparison: Unimodal vs. Multimodal Sentiment Analysis Techniques

The above comparison shows that although every unimodal technique has its strengths, text-based analysis generally has efficiency at its top, then followed by visual and audio methods. Results like that explain why multimodal approaches could become potentially very useful to achieve a more comprehensive sentiment analysis.

11. CONCLUSION AND FUTURE SCOPE OF WORK

Sentiment analysis continues to evolve, providing fertile ground for future research and innovation in several key areas. Below are some of the main directions where advancements are expected in the coming years:

- **Improving multimodal sentiment analysis:** As social media platforms increasingly combine text, images, videos, and emojis in user-generated content, there is a growing need for models capable of interpreting sentiments across multiple modalities. Multimodal sentiment analysis aims to integrate these diverse data types—text, audio, and visual elements—to better understand the full spectrum of emotions expressed online. However, analysing such diverse inputs simultaneously remains a complex task that demands further research. Developing models that can seamlessly combine these data types will improve the accuracy of sentiment analysis, especially on platforms where users' express emotions using more than just text (Khedr et al., 2017).

- **Domain-specific and cross-domain sentiment analysis:** One of the biggest challenges that the current field of sentiment analysis faces is its domain dependence. When we train a model on one dataset or domain-let's say, movie reviews-it performs terribly on another: financial news, because their language, expressions of sentiment, and context are pretty different. The next wave of research will center on cross-domain sentiment analysis, where models need to generalize well across various domains without losing accuracy. Techniques like transfer learning and domain adaptation would, in fact be the key to leveraging generalization of sentiment analysis models and could help them perform well under vastly different conditions (Raghunathan and Saravanakumar, 2023).
- **Real-time sentiment analysis:** Large volumes of data are being created on social media platforms such as Twitter, especially around real-time events such as elections or product launches, and that's exactly why the need for real-time sentiment analysis is increasing. This involves processing large volumes of data while they are being generated, providing actionable insight in near real time. What remains an open challenge is the development of efficient and scalable algorithms with low overhead, capable of handling the dynamic flow of data without compromise on performance or accuracy. Future research will be on ways that further optimize the computational performance and scalability of sentiment analysis models for increased speed and greater reliability in real-time applications (Deveikyte et al., 2020).
- **Handling sarcasm, irony and ambiguity:** Determination of sarcasm, irony, and ambiguous words is most difficult tasks in the area of sentiment analysis. Their meanings are usually indirect and opposite of what their literal meaning indicates, so that a traditional sentiment analysis process cannot capture these cases. As the style of online language continues to evolve and becomes as informal as very informal platforms, such as Twitter, this need comes to the fore more strongly. Future directions may include a BERT model, or a GPT, in which the language and context may provide better mastery in detecting sarcasm or ambiguous sentiments (Chavan et al., 2018).
- **Cross-lingual sentiment analysis:** With the amplification of sentiment analysis around the world, many languages become a challenge to efficiently manage. Most current models of sentiment analysis are developed for only English, though people nowadays demand models that could handle all kinds of languages. In order to address this, cross-lingual sentiment analysis tries to use techniques such as machine translation or pre-trained multilingual models like mBERT in performing sentiment analysis across languages based on less extensive language-specific data. Thus, this research direction is critical to the needs of global businesses and users (Pathak and Rai, 2023).
- **Ethics and sentiment analysis:** The more entrenched sentiment analysis becomes in decision-making processes, the more inextricably linked to ethical concerns will be. Protecting user privacy and security data concerning personal data extracted from social media or other web applications will be absolutely critical. Finally, the question of bias of models as a result of training data bias is directly related to the fairness of sentiments derived from predictions. Future work will be in applying careful development, guidelines, and ethical frameworks for sentiment analysis to the models presented, making sure they are transparent, unbiased, and non-intrusive on users' privacy (Attanasio et. al, 2019).

As seen here, such a field as sentiment analysis leaves much room for growth, and researchers tackling such challenges are part of the way to enable further advancements in this area of research. The future direction of sentiment analysis shall be shaped in the form of advances in multimodal analysis, domain generalization, real-time processing, and ethics.

REFERENCES

[1] Agarwal B, Mittal N, Bansal P, Garg S. Sentiment analysis using machine learning: A survey. Soc Netw Anal Min. 2020;10(1):1-23. doi:10.1007/s13278-020-00678-6.

[2] Alaei A, Becken S, Stantic B. Sentiment analysis in tourism: Capitalizing on big data. J Travel Res. 2019;58(2):175-91. doi:10.1177/0047287517747753.

[3] Appel O, Chiclana F, Carter J, Fujita H. Hybrid models for sentiment analysis. Appl Soft Comput. 2016;47:110-25. doi:10.1016/j.asoc.2016.05.011.

[4] Attanasio G, Salvatore F, Valente A, Porfirione C. Ethics in AI: Trust, bias and responsibility. Artif Intell Rev. 2019;52(4):3119-40. doi:10.1007/s10462-018-9666-2.

[5] Cambria E, Das D, Bandyopadhyay S, Feraco A. Affective Computing and Sentiment Analysis: Emotional States and Personality in Human-Computer Interaction. Springer International Publishing; 2017.

[6] Chandrasekaran G, Nguyen TN. Multimodal sentimental analysis for social media applications. Wiley Interdiscip Rev Data Min Knowl Discov. 2021;11(3). doi:10.1002/widm.1405.

[7] Chavan VS, Nikhade SM, Waghmare A. Detection of sarcasm in text data using machine learning and natural language processing techniques. Int J Comput Appl. 2018;182(47):20-4.

[8] Chopra P, Bhatia A. Sentiment analysis and challenges: A comprehensive study. J Inf Technol. 2016;1(3):45-52.

[9] Consoli D, Palumbo F, Tagarelli A. Forecasting financial markets with sentiment analysis: An overview of existing techniques. IEEE Access. 2020;8:110761-80. doi:10.1109/ACCESS.2020.3001048.

[10] Das R, Singh TD. Multimodal sentiment analysis: Methods and challenges. ACM Comput Surv. 2023;55(4):1-35. doi:10.1145/3548696.

[11] Deveikyte G, Schivinski B, Dabrowski D. Real-time sentiment analysis: Exploring its application and challenges in social media. J Bus Res. 2020;120:294-306. doi:10.1016/j.jbusres.2020.06.048.

[12] Do HH, Prasad P, Maag A, Alsadoon A. Deep learning for aspect-based sentiment analysis: A comparative review. Expert Syst Appl. 2019;118:272-99. doi:10.1016/j.eswa.2018.10.003.

[13] Dong M, Liu B. Financial sentiment analysis: A survey of methods and challenges. ACM Trans Intell Syst Technol (TIST). 2021;12(3):34. doi:10.1145/3455822.

[14] Ebrahimi J, Dou Z, Lowd D, Dou D. Chain-based RNN for relation extraction. Proc 55th Annu Meet Assoc Comput Linguist (Volume 1: Long Papers). 2017;1516-26. doi:10.18653/v1/P17-1139.

[15] Gandhi A, Adhvaryu K, Cambria E. Multimodal sentiment analysis: Systematic review. Inf Fusion. 2023;96:67-78. doi:10.1016/j.inffus.2023.01.014.

[16] Govindarajan S. A comparative analysis of lexicon-based sentiment analysis approaches. J Data Sci Anal. 2022;10(2):132-45.

[17] Islam MR, Zibran MF. Sentiment analysis of open-source software repositories: A systematic literature review. Inf Softw Technol. 2018;102:81-91. doi:10.1016/j.infsof.2018.05.005.

[18] Jefferson A, Roberts S. Fuzzy logic and rule-based sentiment analysis approaches: A review. J Comput Linguist. 2017;43(4):345-60.

[19] John D, Kartheeban S. A comprehensive study on k-fold cross-validation in machine learning. J Artif Intell Res. 2019;6(2):52-60.

[20] Kaur R, Kautish S. Multimodal sentiment analysis: A survey and comparison. In: Kautish S, editor. Handbook of Research on Machine Learning Techniques for Pattern Recognition and Information Security. Hershey, PA: IGI Global; 2022. p. 250-70.

[21] Khedr AM, Hefny H, Ramadan ZH. Multimodal sentiment analysis using deep learning for social media applications. Int J Adv Comput Sci Appl. 2017;8(5):30-6. doi:10.14569/IJACSA.2017.080505.

[22] Köksal M, Özgür A. Turkish sentiment analysis: Challenges and approaches. J Inf Sci. 2021;47(4):532-45. doi:10.1177/0165551520912021.

[23] Kumaresan N, Thangaraju A. Challenges in multimodal sentiment analysis: A critical survey. ACM Comput Surv (CSUR). 2023;55(1):14. doi:10.1145/3432934.

[24] Lyu X, Liu X, Liang W. Political sentiment analysis: Insights from computational social science. J Comput Soc Sci. 2023;6(3):421-45. doi:10.1007/s42001-022-00171-2.

[25] Mahersia H, Ezzameli K. Emotion recognition from unimodal to multimodal analysis: A review. Pattern Recognit Lett. 2023;169:1-12. doi:10.1016/j.patrec.2023.05.017.

[26] Maiya V, Patil S. Exploring multimodal sentiment analysis for enhancing customer experience. J Bus Res. 2020;121:561-72.

[27] Mäntylä M, Graziotin D, Kuutila M. The evolution of sentiment analysis: A review of research trends and challenges. J Comput Soc Sci. 2016;3(1):199-224.

[28] Mohammad S. Challenges of sarcasm detection in sentiment analysis. J Comput Linguist. 2017;43(2):1-27. doi:10.1162/COLI_a_00296.

[29] Nabil M, Aly M, Atiya A. LABR: A large scale Arabic book reviews dataset. Proc Conf Empir Methods Nat Lang Process (EMNLP). 2014;1-11. doi:10.3115/v1/D14-1097.

[30] Nabil M, Aly M, Atiya A. ASTD: Arabic sentiment tweets dataset. Proc 8th Int Conf Lang Resour Eval (LREC). 2015;4002-9.
[31] Nazar K, Bhattasali T. Precision, recall, F1 score for sentiment analysis: A review. J Data Sci Mach Learn. 2021;3(2):102-10.
[32] Ovrelid L, Velldal E, Bergem T. NoReC_fine: A fine-grained sentiment analysis dataset for Norwegian. Proc Conf Empir Methods Nat Lang Process (EMNLP-IJCNLP). 2019;6350-7.
[33] Poria S, Cambria E, Bajpai R, Hussain A. A review of affective computing: From unimodal analysis to multimodal fusion. Inf Fusion. 2019;37:98-125.
[34] Poria S, Cambria E, Hazarika D, Vij P. Multimodal sentiment analysis: Addressing key issues and setting up the future. IEEE Intell Syst. 2020;35(6):17-25. doi:10.1109/MIS.2020.2985612.
[35] Pathak P, Rai P. Cross-lingual sentiment analysis using multilingual transformer models: A survey. ACM Trans Asian Low-Resour Lang Inf Process (TALLIP). 2023;22(1):12. doi:10.1145/3484842.
[36] Potts A, Brown S, Evans P. Cohen's kappa and the measurement of agreement in sentiment analysis: An exploration. J Comput Soc Sci. 2020;3(1):199-224. doi:10.1007/s42001-019-00063-9.
[37] Pozzi FA, Fersini E, Messina E, Liu B. Cross-lingual sentiment analysis: A review and future directions. ACM Trans Inf Syst (TOIS). 2017;36(1):12. doi:10.1145/3119927.
[38] Quan C, Ren F. On improving the accuracy of product recommendation systems with sentiment analysis. J Comput Linguist Nat Lang Process. 2016;3(5):100-10.
[39] Raghunathan S, Saravanakumar R. Domain-adaptive sentiment analysis: Challenges and future trends. Expert Syst Appl. 2023;236:120444. doi:10.1016/j.eswa.2022.120444.
[40] Ribeiro FN, Araujo M, Gonçalves P, Andre G, Benevenuto F. Sentibench: A benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci. 2015;4(1):23. doi:10.1140/epjds/s13688-015-0057-1.
[41] Sebastiani F. Sentiment quantification: Estimating the prevalence of sentiment classes in social media posts. IEEE Intell Syst. 2016;31(6):62-7. doi:10.1109/MIS.2016.68.
[42] Shayaa S, Hassan SU, Othman M. Tackling data imbalance and sparsity in sentiment analysis: Current strategies and future directions. Data Sci Anal. 2018;7(2):45-60. doi:10.1007/s41060-018-0073-y.
[43] Soleymani M, Garcia D, Jou B, Schuller B, Chang S-F, Pantic M. A survey of multimodal sentiment analysis. Image Vis Comput. 2017;65:3-14.
[44] Sun L, Lian Z, Liu B. Efficient multimodal transformer for sentiment analysis. IEEE Trans Neural Netw Learn Syst. 2023;34(8):3456-67. doi:10.1109/TNNLS.2023.3256894.
[45] Tripathi S, Bagui S, Palit P. Sentiment analysis on IMDb movie reviews dataset using hybrid approaches. J Big Data. 2020;7(1):55. doi:10.1186/s40537-020-00317-4.
[46] Valverde-Albacete FJ, Pelayo F, Villarroel J. Entropy-based evaluation metrics for sentiment classification. Inf Sci. 2013;250:230-47. doi:10.1016/j.ins.2013.07.011.
[47] Wu L, Guo H, Shi S, Yang X. Product recommendation systems using sentiment analysis of user reviews. Knowl-Based Syst. 2017;131:23-34. doi:10.1016/j.knosys.2017.06.034.
[48] Xu H. Multi-Modal Sentiment Analysis. In: Xu H, editor. Multimodal Sentiment Analysis: Theory and Application. Singapore: Springer; 2023. p. 123-45. doi:10.1007/978-981-99-5776-7_6.
[49] Zadeh A, Chen M, Poria S, Cambria E, Morency L-P. Multimodal sentiment intensity analysis in videos: Facial, verbal, and vocal fusion. IEEE Intell Syst. 2018;34(2):52-8.
[50] Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. Wiley Interdiscip Rev Data Min Knowl Discov. 2018;8(4)
[51] Zhang C, Zhu L, Kong X. Multimodal sentiment analysis based on fusion methods: A survey. Inf Fusion. 2023;95:306-25. doi:10.1016/j.inffus.2023.03.004.
[52] Zhou Y, Ye Y. Applications of sentiment analysis in education. Educ Technol Soc. 2020;23(1):115-27.