IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

Multilinguistic Sentimental Analysis using AWS

Dr. Ch. Vijaya Kumar[1], Madhagoni Sravani[2], Ravula Sandhya[3], Pavani K M[4], Bairu Shiva Kumar[5]
Professor & HOS, Computer Science Engineering[1]
Students, Computer Science Engineering[2,3,4,5]
ACE Engineering College, Ghatkesar, India

Abstract: *Multilingual sentiment analysis is a technique that cracks the code of emotions across languages. It analyses written text, like social media posts or customer reviews, to understand if the overall tone is positive, negative, or neutral. This is particularly useful for businesses that operate internationally, as it allows them to gather insights from a wider audience that speaks various languages. Multilingual sentiment analysis is a technique that cracks the code of emotions across languages. It analyses written text, like social media posts or customer reviews, to understand if the overall tone is positive, negative, or neutral. This is particularly useful for businesses that operate internationally, as it allows them to gather insights from a wider audience that speaks various languages. This paper reviews the use of Amazon Web Services (AWS) for multilingual sentiment analysis, exploring the underlying technologies, advantages, challenges, and performance benchmarks. We discuss the architecture for implementing such solutions and examine real-world applications. The paper highlights how AWS's cloud services simplify the deployment of scalable and efficient sentiment analysis models, making them accessible for a global audience.*

Keywords: Multilinguistic sentimental analysis, Amazon comprehend, Language detection, Customer feedback analysis, Amazon Web Services

I. INTRODUCTION

Multilingual sentiment analysis allows us to analyze sentiments in other languages that facilitate world communication. The conventional methods are not so effective for other languages, but cloud environments such as AWS offer simple and real-time processing. The present project based on AWS services is centered on improving sentiment recognition in other languages by using machine learning models. Overcoming different challenges languages provide guarantees precise sentiment classification through NLP methods. The present research is centered on efficient automized sentiment analysis for the broader linguistic geography; outside the scope of carried-out studies.

[1] This system concerns certain of the methods of sentiment analysis of text data using machine learning and deep learning. It compares test metrics, model architectures, and training procedures on multiple datasets. The findings indicate the capability of neural networks in capturing subtle sentiment.

[2] In this work, the authors address sentiment analysis of syntactically and semantically rich code-mixed languages, i.e., Roman Urdu-English tweets. They employ state of the art transformer models Electra, cm-BERT, and m BART to improve sentiment prediction performance on multilingual and informal social media datasets. The work also involves topic modeling through LDA to identify common themes and trends in sentiment clusters. The work shows the promise of fine-tuned multilingual transformers in low-resource, linguistically rich settings.

[3] A two-modality approach is proposed for sentiment analysis of both spoken audio and their transcripts. it combines signal processing and natural language techniques for overall sentiment detection. Experimental results demonstrate enhanced prediction performance with multimodal fusion.

[4] This system presents a new multimodal dataset including video, audio and text inputs for sentiment analysis of car reviews. It describes the data acquisition process and proposes approaches for multimodal feature extraction. The results set the superiority of cross-modal cues for sentiment comprehension.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 9, May 2025 Impact Factor: 7.67

[5] The system integrates a wide range of sentiment analysis methods from various fields. It classifies methods and shows their applicability in practice and technological advancements. Some of the key challenges like detecting sarcasm and domain adaptability are also addressed.

[6] This system classifies the methods of sentiment analysis into rule-based, statistical, and neural network methods. It talks about training and performance measurement using annotated datasets. Future research directions include hybrid modeling and cross-domain transfer learning.

[7] This system proposes a multi-task learning solution to identify student sentiment from course feedback. Selective paraphrasing is used to generalize statements of feeling in comparable situations. The approach realizes aspect detection and opinion polarity classification enhancements.

[8] The article seems to be talking about language diversity and multilingual education policy in Europe. It likely investigates linguistic integration, pedagogy, and sociopolitical concerns. Because of encoding problems, the rich content could not be recovered.

[9] The article thoroughly discusses sentiment analysis methods, i.e., lexicon, rule-based, and machine learning. It breaks down performance metrics and data sources by industry into various categories such as social media, product reviews, and politics. All the tasks like sarcasm detection, domain adaptation, and multilingual processing are framed as open research tasks.

[10] This article proposes a sentiment analysis model that integrates both machine learning and lexicon-based methods. It introduces a new genetic algorithm-based feature selection approach that is more scalable and accurate for classification. Experimental outcomes exhibit up to 42% feature-set reduction and improved accuracy over PCA and LSA approaches on different datasets.

II. LITERATURE SURVEY

This project explores multilingual sentiment analysis, with an emphasis on cloud-based service integration such as Amazon Web Services (AWS) for scalable and efficient sentiment extraction across multiple languages. As global communication cuts across linguistic boundaries, the need for strong sentiment analysis software supporting multiple languages has increased. Amazon Web services such as Amazon Comprehend and its multi-language capability provide the feasibility of real-world application. This paper provides a review of five recent studies between 2020 and 2025 that tackle multilingual sentiment analysis from different perspectives, such as cloud deployment, zero-shot learning, and transformer models. The objective is to identify current methods, performance, and limitations of multilingual sentiment detection. This review also approximates the ease of real-world implementation with AWS.

[11] This system provides a comparative assessment of AWS and Microsoft Azure cloud-based sentiment analysis APIs. The authors developed an application deployable on both platforms, with performance as quantified by common parameters of accuracy, precision, and recall. It concludes that Amazon Comprehend is likely to provide better sentiment detection, particularly for multilingual data sets. The paper provides a good approach to testing commercial sentiment APIs, with ease of integration and use being a priority. It acknowledges AWS's expertise in text processing rich in context, although acknowledges limitations in language-specific sentiment subtleties. The practical approach of this paper is particularly helpful for developers. It also acknowledges some cost implications in scaling services. Overall, this research confirms that AWS provides good multilingual sentiment detection for practical use.

[12] This work investigates the application of transformer models such as XLM-Roberta for zero-shot multilingual sentiment analysis. It evaluates performance across many languages without explicit training, relying on pretrained multilingual embeddings. The result shows promising accuracy, particularly on high-resource languages, but also highlights challenges with sentiment detection on low-resource ones. The work suggests the possibility of using such models with cloud platforms such as AWS for scalable deployment. Although AWS is not the focus, the method is readily transferable to platforms such as Amazon SageMaker. The result of the work affirms zero-shot learning as an effective method to multilingual sentiment analysis, particularly in real-time, dynamic settings. It further calls for enhanced cross-lingual data augmentation. Overall, this work is in the midst of espousing robust sentiment analysis across low-resource languages.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

[13] This system seeks to leverage machine translation (MT) to address sentiment analysis in low-resource languages. The authors employ English sentiment models and employ MT to translate target language input text. With some translation noise, the method produces workable sentiment outputs. The solution can be scaled using cloud translation APIs and AWS Comprehend for processing. It shows that translation-based pipelines can be a low-cost and workable solution where native sentiment models are unavailable. The research identifies limitations of losing sentiment-specific cues during translation. It sets the stage for a hybrid MT-sentiment model pipeline that is well adapted to cloud deployment scenarios. The method has potential to be incorporated within multilingual sentiment tools used in global applications.

[14] This system assesses the contribution of instruction-tuned large language models (LLMs) to multilingual sentiment analysis. By providing task instructions to models such as mT5 and LLaMA variants, the research attains state-of-the-art performance across several languages. It proves the success of prompt engineering and fine-tuning methods in achieving accuracy and processing code-switching. The research also prioritizes reproducibility and presents benchmark outcomes on standard datasets. Even though the paper does not use AWS directly, its models can be hosted using AWS services such as SageMaker or EC2 instances. This versatility affirms AWS's status as a backbone for hosting LLMs. The paper concludes that instruction-tuned LLMs are ideally positioned for multilingual NLP tasks. It also points to additional opportunities for domain adaptation.

[15] This paper investigates cross-lingual transfer learning using pretrained models like XLM and mBERT for sentiment analysis. The concept is training on English and testing on foreign languages without fine-tuning. Performance varies with language families, with the Indo-European language family gaining the most from transfer. The method reduces the requirement for annotated multilingual data and is ideally suited to scalable architecture on cloud platforms. Although not AWS-focused, the research methodology fits seamlessly into SageMaker's multilingual NLP pipelines. The paper emphasizes careful model selection and preprocessing for better performance. It also gives a glimpse of the future of universal NLP models in multilingual environments. This paper significantly enhances scalable, cloud-compatible sentiment solutions.

III. PROPOSED WORK

Methodology of Sentimental Analysis using AWS

The flowchart depicts the sequential workflow of a multilingual sentiment analysis project based on AWS services. The process starts with creating an AWS account and IAM role setup to provide secure access management. YouTube API is employed to retrieve comment datasets, which are uploaded to Amazon S3 buckets for storage. The data stored is processed to retrieve output files having sentiment outcomes. Sentiment scores are then derived, giving insights about user opinions in various languages.

B. AWS Account Setup and IAM Role Configuration

To start the project, an AWS root user account was established, which provided access to a wide variety of cloud services provided by Amazon Web Services. Security measures were adopted by creating an IAM (Identity and Access Management) role rather than utilizing root access for operations. The IAM role was set up with particular permissions for Amazon Comprehend, Amazon S3, and IAM policies, providing limited but adequate access for sentiment analysis and for data storage. The permissions to Amazon Comprehend were added to provide access to language detection and sentiment analysis capabilities. S3 access permissions were added to control buckets and store input-output files. The IAM role was pivotal in securely handling resources while allowing operational boundaries. Correct IAM configuration minimized the risk of unwanted access and improved automation of sentiment analysis tasks. The configuration was mandatory to allow the project to scale, be secure, and conform to AWS best practices. The IAM role's access credentials were utilized during code execution and during the creation of analysis jobs. Having backend permissions configured in place, more development went towards data collection and processing.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

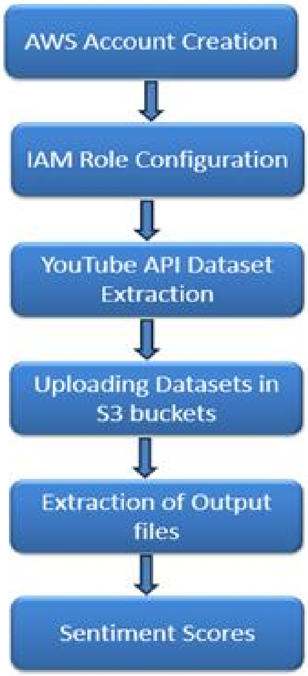


Fig. 1. Workflow of the sentimental analysis

C. Dataset Preparation and S3 Bucket Configuration

After configuring the AWS environment, Amazon S3 (Simple Storage Service) was used to store multilingual datasets. Separate S3 buckets were created, and each bucket contained organized folders named input and output for better data management. Python was used to extract user comments from YouTube using the YouTube Data API, focusing on gathering a diverse set of comments in English, Korean, French, and Japanese. Each language dataset was saved as an individual CSV file to maintain clarity during processing and analysis. The CSV files were uploaded to the input folder of the S3 bucket for further sentiment analysis. The CSV structure included serial numbers, comment text, language codes, and metadata such as like counts. Care was taken to clean the data and remove noise such as emojis, special characters, and spam-like entries before uploading. This step ensured that the input to Comprehend was clean and would yield meaningful sentiment scores. With organized and language-specific datasets in place, the next step involved configuring Amazon Comprehend for analysis. The S3 bucket structure simplified the integration with Comprehend jobs by clearly specifying input and output data paths.

D. Sentimental Analysis using AWS Comprehend

The central component of the project was to apply Amazon Comprehend to carry out sentiment analysis on the uploaded multilingual data sets. For every language, a distinct analysis job was submitted through the Amazon Comprehend console. The job was set up by indicating the input and output folder paths in the respective S3 bucket. Amazon Comprehend processed the uploaded CSV files and classified the sentiments into four categories: Positive, Negative, Neutral, and Mixed. At job completion, the output was saved in the output folder in compressed .tar.gz formats per language. The files contained structured JSON-like sentiment data per comment. Comprehend automatically managed language detection and analysis without needing a distinct model per language, making the workflow streamlined.The analysis gave in-depth insights into user opinions of various linguistic origins. The capability to process multiple languages in one architecture highlighted the strength of AWS Comprehend. After successfully completing the jobs, the results were ready to be extracted and processed further. This facilitated moving to the next stage, where results were extracted and transformed into a form that could be used.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 9, May 2025

Impact Factor: 7.67

Stage	Description	Tools Used
Data Collection	Collecting the multilingual text data using YouTube API	YouTube API
Language Detection	Identifying language of the input text	AWS Comprehend
Sentimental Analysis	Categorize text as positive, Negative, and Neutral	AWS Comprehend
Entity Recognition	Identify keywords and concepts from the text	AWS Comprehend
Data Storage	Storing processed sentimental data securely	AWS S3

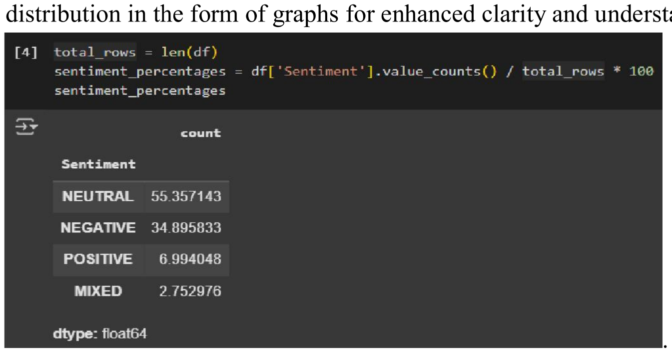
Table I. Tools used in the work flow of sentiment

E. Post-Processing, Categorization and Sentiment Score Calculation.

After downloading the .tar.gz files from the S3 output directory, they were unpacked to yield sentiment analysis reports. Each of the archives held structured data files of individual comments and their respective sentiment category and confidence scores. A Java application was authored for reading the extracted files automatically, classifying them as positive, negative, and neutral comments, and converting them into CSV format. This conversion facilitated smoother integration with visualization and aggregation tools. Python was later employed to calculate the overall sentiment scores for every language after generating the categorized CSVs. The sentiment counts were aggregated and normalized to determine the percentage of every sentiment category. This scoring assisted in the ability to know which languages contained more positive or negative user opinions through comment analysis. Overall results were represented visually through the use of libraries like Matplotlib and Seaborn to produce bar charts and sentiment distribution plots. Post-processing was critical in converting raw data from analysis into actionable insights. The end-to-end pipeline supported multilingual sentiment extraction, processing, and visualization through a mix of AWS services and custom code.

IV. RESULTS

These findings present the output of our multilingual sentiment analysis system for two languages. Through AWS Comprehend, every comment from the uploaded data is identified automatically for its language and subsequently analysed to find its sentiment whether positive, negative, neutral, or mixed along with accuracy confidence. The processed data is saved and fetched by using YouTube API and is displayed in the form of structured tables. We have also represented sentiment distribution in the form of graphs for enhanced clarity and understanding.



```
[4] total_rows = len(df)
    sentiment_percentages = df['Sentiment'].value_counts() / total_rows * 100
    sentiment_percentages
```

Sentiment	count
NEUTRAL	55.357143
NEGATIVE	34.895833
POSITIVE	6.994048
MIXED	2.752976

dtype: float64

Fig. 3. 1 Overall Sentimental Score of English Comments

The first language heatmap graphically represents the intensity and distribution of sentiments Positive, Negative, Neutral, and Mixed throughout the whole dataset. Every row in the heatmap is a single comment, and the intensity of the color is a measure of the confidence score for each sentiment. A darker color under the column of "Positive," for instance, indicates the system has a strong confidence that the comment holds a positive sentiment. This visual depiction aids in identifying patterns in a quick and easy manner, like whether the English comments are majority biased towards a certain sentiment or evenly distributed. It aids in the detection of outliers or mixed sentiments where no single sentiment dominates. The heatmap can be utilized for interpreting large quantities of data in a glance and making decisions based on that.

DOI: 10.48175/IJARSCT-27067

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

```python
from matplotlib import pyplot as plt
import seaborn as sns
import pandas as pd
plt.subplots(figsize=(8, 8))
df_2dhist = pd.DataFrame({
    x_label: grp['Sentiment'].value_counts()
    for x_label, grp in df.groupby('File')
})
sns.heatmap(df_2dhist, cmap='viridis')
plt.xlabel('File')
_ = plt.ylabel('Sentiment')
```

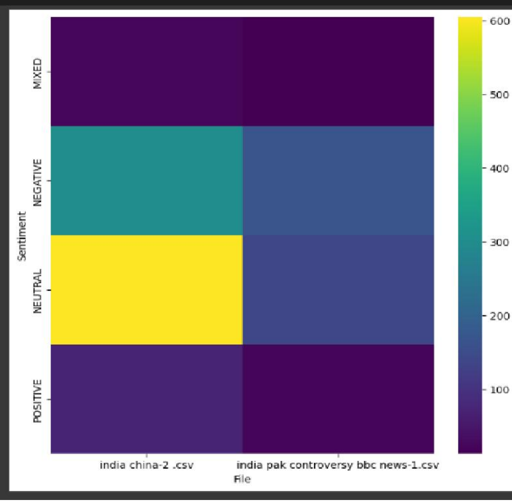


Fig.3.2 Heatmap of Sentimental Analysis of English Comments

The overall sentiment score of English-language comments on YouTube is captured through both quantitative measures and graphical aids for easy interpretation. Quantitatively, the score indicates the overall sentiment polarity e.g. positive, neutral, or negative derived from the analysis of all English language comments. Graphically, this information is most commonly presented in the form of bar charts, line graphs, or heatmaps to show the distribution and strength of sentiments over time or by category.

```python
total_rows = len(df)
sentiment_percentages = df['Sentiment'].value_counts() / total_rows * 100
sentiment_percentages
```

Sentiment	count
NEUTRAL	50.310559
POSITIVE	41.304348
NEGATIVE	8.074534
MIXED	0.310559

dtype: float64

Fig. 3. 3 Overall Sentimental Score of Korean Comments

The images below demonstrate the close-up of the sentiment score and heatmap for Korean comments. The heatmap of strong positive or negative reactions is pointed out by it. This language-agnostic model classifies comments in any language by identifying the language, translation if necessary, and sentiment classification to generate reliable scores.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

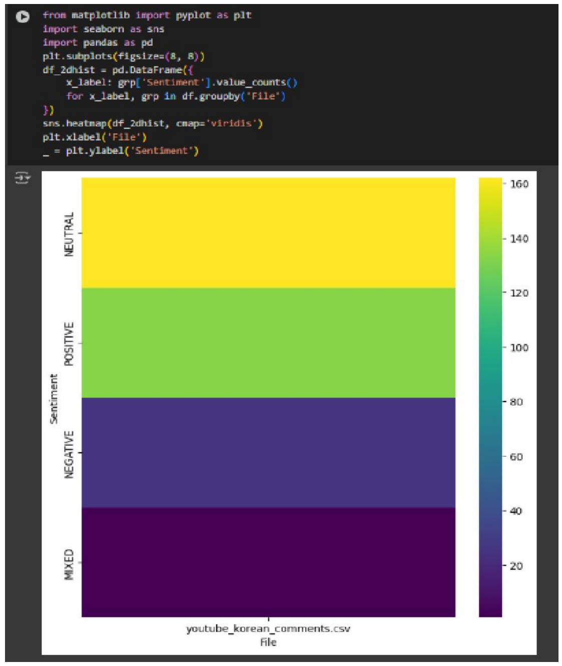


Fig.3.4 Heatmap of Sentimental Analysis of Korean Comments

These outputs includes both numeric sentiment scores and understandable visual representations, allowing comparative analysis in various languages and cultural environments. This multi-language capacity guarantees the wide application of the system across global platforms to enable organizations to comprehend and effectively act on user feedback irrespective of language differences.

V. CONCLUSION

This project proves to be successful in how cloud-based technology and NLP services can be used to do multilingual sentiment analysis on actual data from sites such as YouTube. Through the use of AWS Comprehend and the YouTube Data API, we were able to gather, process, and examine user commentary in various languages, providing useful information on public opinion that can help businesses optimize their marketing efforts, product excellence, and customer satisfaction. Even with the constraint of the AWS Comprehend free-tier data size, we managed to overcome this obstacle using efficient data preprocessing and reduction methods. The system not only identifies the language used in every comment but also properly categorizes the sentiment as positive, negative, neutral, or mixed and is then interpreted using graphs to facilitate easier understanding. On the whole, this multilingual sentiment analysis model offers a scalable and automated answer for deciphering global customer opinions. The system can be further developed in the future with additional languages added, utilizing real-time streaming data, and utilizing dashboards for real-time insights, enabling organizations to make even more data-driven decisions.

VI. ACKNOWLEDGE

We are also very thankful to Dr. Ch. Vijaya Kumar, Professor and HOS, Department of Computer Science Engineering, ACE Engineering College, for his thoughtful guidance, advice, and valuable suggestions all through this project. We also appreciate our institution for the resources and support we received. Above all, we would like to extend our sincere appreciation to the editorial team of IJARSCT for allowing us to publish our work.

REFERENCES

[1] Mamani-Coaquira, Yonatan, and Edwin Villanueva. "A Review on Text Sentiment Analysis With Machine Learning and Deep Learning Techniques." *IEEE Access* (2024).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

[2] Hashmi, Ehtesham, Sule Yildirim Yayilgan, and Sarang Shaikh. "Augmenting sentiment prediction capabilities for code-mixed tweets with multilingual transformers." *Social Network Analysis and Mining* 14.1 (2024): 86.

[3] Dhariwal, Naman, Sri Chander Akunuri, and K. Sharmila Banu. "Audio and Text Sentiment Analysis of Radio Broadcasts." *IEEE Access* 11 (2023): 126900-126916.

[4] Stappen, Lukas, et al. "The multimodal sentiment analysis in car reviews (muse-car) dataset: Collection, insights and improvements." *IEEE Transactions on Affective Computing* 14.2 (2021): 1334-1350.

[5] Wankhade, Mayur, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. "A survey on sentiment analysis methods, applications, and challenges." *Artificial Intelligence Review* 55.7 (2022): 5731-5780.

[6] Tan, Kian Long, Chin Poo Lee, and Kian Ming Lim. "A survey of sentiment analysis: Approaches, datasets, and future research." *Applied Sciences* 13.7 (2023): 4550.

[7] Gul, Shahla, et al. "Advancing Aspect-Based Sentiment Analysis in Course Evaluation: A Multi-Task Learning Framework with Selective Paraphrasing." *IEEE Access* (2025).

[8] Tleuzhanova, G. K. "Multilingual education in the European countries: analysis, experience, problems." *Bulletin of the Karaganda university Pedagogy series* 98.2 (2020): 112-120.

[9] Mao, Yanying, Qun Liu, and Yu Zhang. "Sentiment analysis methods, applications, and challenges: A systematic literature review." *Journal of King Saud University-Computer and Information Sciences* (2024): 102048.

[10] Iqbal, Farkhund, et al. "A hybrid framework for sentiment analysis using genetic algorithm based feature reduction." *IEEE Access* 7 (2019): 14637-14652.

[11] Ivan, Sergiu C., Robert Ş. Győrödi, and Cornelia A. Győrödi. "Sentiment Analysis Using Amazon Web Services and Microsoft Azure." *Big Data and Cognitive Computing* 8.12 (2024): 166.

[12] Anitha, R., and KS Anil Kumar. "Sentiment Analysis in low resource language: exploring Bert, mbret, xlm-r, and RNN architectures to underpin the deep language understanding."

[13] Antony, Allen, et al. "Leveraging multilingual resources for language invariant sentiment analysis." *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. 2020.

[14] Hasan, Md Arid. "Ensemble Language Models for Multilingual Sentiment Analysis." *arXiv preprint arXiv:2403.06060* (2024).

[15] Thakkar, Gaurish, Sherzod Hakimov, and Marko Tadić. "M2sa: multimodal and multilingual model for sentiment analysis of tweets." *arXiv preprint arXiv:2404.01753* (2024).