A FRAMEWORK FOR FINE GRAINED SENTIMENT ANALYSIS ON CODE-MIXED LANGUAGE FOR SOCIAL MEDIA USER BEHAVIOURS

Anand, Tank[1]; Pratik, Vanjara[2]

[1]Atmiya University – Rajkot, India

[2]Shree M. P. Shah Commerce College – Surendranagar, India

ABSTRACT

Here, we provide a framework that discovers sentiments from social media platforms, assesses, and transforms them into meaningful data. Social media is changing people's attitudes and habits, which in turn is influencing their choices. Attempting to keep an eye on social networking activity is a useful tool for tracking consumer attitude about products and firms and gauging loyalty from consumers. The framework can be the next natural area for branding based on the polarities on the internet and social media. We present a dynamic solution method for sentiment analysis using the classification of interpersonal data sources. To evaluate the caliber of social information services, we also introduce a brand-new quality model. We utilize public comments, posts through social media as an inspiring case study. Specifically, to pinpoint the comments and posts we concentrate on the spatiotemporal characteristics of the attitudes expressed by social media users. On datasets from the real world, experiments are carried out. Our suggested model's performance is preliminary demonstrated by performance evaluation matrix.

Keywords – Sentiment Analysis, social media, Artificial Intelligence, Machine Learning, Natural Language Processing

1. INTRODUCTION

Societies use their sentiments to convey their opinions and behaves as per experience. Sentiment analysis is the term used to describe the analysis of these kinds of perspectives. Social networks, or social networking sites, like Facebook, Twitter, and others, have become a free community for dataset [1]. When there is an event, a lot of people use social media services to quickly create and share data i.e. Twitter, Facebook, LinkedIn, etc. [2]. This data generated by social media users has many useful information as metadata: 1) It has subjective information as Opinion. 2) It includes spatiotemporal information as Behavior, among others.

Extraction and comprehension of human dynamics, including actions, trends, attitudes, and emotions, are made easier by sentiment analysis of subjective data [3]. Through forums, blogs, wikis, social networks, and other online tools, millions of people communicate their thoughts and feelings. Sentiment Analysis (SA) examines a user's thoughts, attitudes, perspectives, ideas, beliefs, remarks, requests, inquiries, and preferences toward various things such as services, issues, people, products, events, subjects, organizations, and their characteristics, based more on emotion than on logic in the form of writing. It determines the general mood of the writer for a text, which could include speeches, product evaluations, blog entries, online forums, database sources, social media information, and papers. Depending on the circumstances, it typically consists of three components: 1) Opinions or emotions, 2) Subject, 3) Opinion Holder [4].

Sentiment analysis of subjective datasets facilitates the extraction and understanding of human dynamics, including behaviors, patterns, attitudes, and emotions [5]. A lot of noise, or superfluous and useless material, is frequently present in social Datasets. Furthermore, there are many kinds of social media information services available on the internet. On the internet, there are numerous data providers that have distinct data characteristics, such as size, quality, and so on. Thus, every dataset needs a different approach for extracting useful information. It takes time to use different tools, and the perspectives of the data from social sensors are not always consistent. [6].

Our study presents a system called "Fine Grained Sentiment Analysis for Code-Mixed Languages" (FGSACML), which gathers sentiments for code-mixed language from multiple social media sites, analyzes and transforms them into meaningful data, and then returns the data in the form of a polarity. Using different characteristics of the data from the social media user, we analyzed social information like Positive, Negative and Neutral. Conventional techniques for sentiment evaluation also cover the numerous attributes of media information like comments, posts and review. Nevertheless, this framework transcends polarities and is not just based on sentiment. Two primary methods can be used

to accomplish this classification. 1) Machine Learning Approach; and 2) Lexicon Based Approach.

The rest of the manuscript is structured as follows. Part II outlines SA methods. Section III contains the literature review. Section IV offers a summary of the framework, followed by a comprehensive report in Section V, which also includes suggestions for future research.

2. METHODOLOGY

Sentiment analysis is an automatic method of determining if a text is created or utilized to represent a consensus, a negative, or a satisfactory opinion of the subject matter. There are three levels of the sentiment classification: Document classification, Sentence classification, Aspect classification, and Feature classification [7].

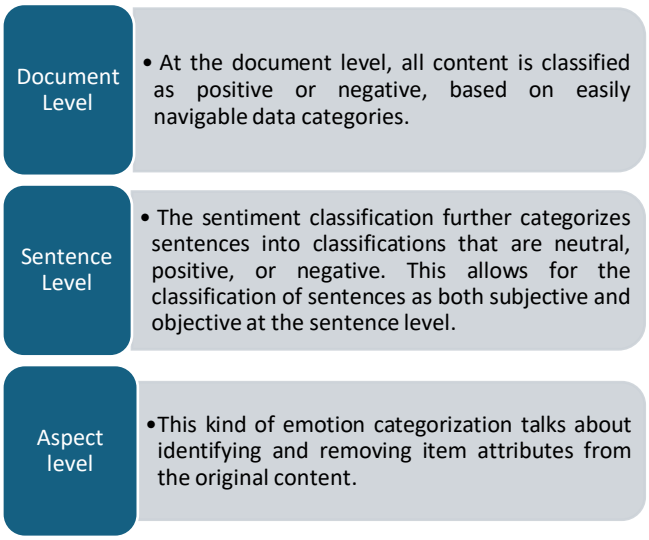

Figure 1 – Classification Level of Sentiment Analysis

NLP is the study of Computer-Assisted Speech and language Processing (CASP) and its application. NLP is extensively utilized in automated inquiry, machine interpretation, and text mining. Previous studies shows that NLP uses a wide range of techniques in different domains and the results are above 80%, which can be said to be quite impressive in comparison to Machine Learning [8]. For SA, the following four classification approaches are implied: 1) Machine Learning Method 2) Lexicon-Based Method 3) Hybrid Method 4) An alternative method.

2.1 Machine Learning Approach

There are two main categories for machine learning models. 1) Supervised Learning 2) Unsupervised Learning. Supervised learning is the highest applied machine learning technique [8]. Using labeled source data, a model is trained using this technique. When fresh unlabeled input data is obtained, the trained model can forecast an output. Supervised learning is typically more effective than unsupervised or semi-supervised learning techniques. However, depending only on labeled training data might be time-consuming and ineffective [9]. Random Forests, Support Vector Machines, Logistic Regression, Naive Bayes, and Decision Trees are among the classifiers applied in sentiment analysis [10].

Machine learning permits computers to pick up new abilities without requiring additional programming. Popular techniques include Decision Trees (DT), Logistic Regression (LR), K-nearest Neighbors (KNN), Support Vector Machines (SVM), Naïve Bayes (NB), Maximum Entropy (ME), semi-supervised and super-supervised learning.[11]. This data leads them to the conclusion that, in all cases, Logistic Regression outperforms the other classifiers, SVM and LR based classifiers perform well, while DT based classifiers provide superior accuracy [12]. Specifically, several academics have examined CNN and its various CNN+LSTM combination. Their performance was contrasted with that of modern models, including CNN/LSTM, KNN/NN, etc. With an accuracy rate of over 97%, this model outperforms all other models. In the CNN LSTM (F-Measure-88, Mean-91, and Probability-96.32%) fare better than the other models (F-Measures-92, Mean-93, and Probability-97.8%). Their evaluation results demonstrate the superiority of the approach over all deep learning models, with BRCAN scoring the highest at 96.32% [13].

2.2 Lexicon Based Approach

The strategy based on Lexicons isolates the words when applying lexicon-based strategies that are accessible for a specific text. Usually, scores are added together to achieve this. Essentially, it is split into two sections: a) Dictionary-based b) corpus-based [11]. Words that can be used in conjunction with sentiment words, such as increment, decimals, and inverted words, were not counted in the previous uni-gram lexicon system; only sentiment words were [12]. SVM and lexicon-based classification are the two methods used for analysis at the aspect level. Lexicon-based models are 84% less accurate than SVM [14]. Sentiment analysis may be easily performed at the feature and sentence level thanks to the lexicon-based approach. It might be considered an unsupervised method because the training data doesn't need to be processed [11]. Lexicon sentiment analysis, commonly referred to as "Dictionary Sentiment Analysis," is the process of analyzing an article's sentiment using a predetermined list of terms and their sentiment ratings. The method is based on a lexicon or dictionary of words along with their polarity—that is, whether a term has a “positive”, “negative”, or “neutral” meaning. One advantage of the Lexicon-based method is that it doesn't expect any training data, likewise some experts even refer to it as an unsupervised approach [15]. The primary drawback of the lexicon-based method is its strong domain specialization, which prevents terms from one domain from being utilized in another [16].

3. LITERATURE REVIEW

Table 1 – Analysis of SA methods for Indian Languages

Ref.	Language	Classification Method Used	Level of Accuracy
[17]	HINDI	SVM	78.14
[18]		Naïve Bayes	87.1
[19]		Lexicon Based	70%
[20]	Gujarati	Guj-Sento Word Net, Bag-of-Word, Word Net	52.27%
[21]		SVM	92%
[22]	Bengali	SVM	98.7%
[23]	Punjabi	Naïve Bayes	Not Available
[24]		Decision Tree	Not Available
[25]	Tamil	SVM	75.9%
[26]	Kannada	Decision Tree	79%
[27]	Malayalam	SVM	91%
[28]		Lexicon Based	85%

Table 1 summarizes SA for classification methods and their accuracies across different datasets based on Indian language. In Hindi sentiment analysis, Naive Bayes achieved 87.1% accuracy [18], while SVM reached 78.14% [17]. Gujarati tweet classification saw SVM outperforming lexicon-based methods, with accuracies of 92% [21] and 52.27% [20]. Bengali sentiment analysis with SVM attained an impressive 98.7% accuracy [22]. Tamil sentiment analysis using SVM achieved 75.9% accuracy [25], and Kannada sentiment analysis with Decision Tree reached 79% accuracy [26]. Malayalam sentiment analysis showed SVM's accuracy at 91 % [27] and Lexicon Based approach at 85% [28]. Hybrid Machine Learning, particularly SVM, demonstrated the highest average precision of 82.9% [29].

Table 2. Performance Analysis for SA Techniques

Techniques	Accuracy	Ref.
Naïve Bayes, SVM, Maximum Entropy	81.0%, 81.5%, 82.9%	[30]
Naive Bayes with Subject identification	87.2%	[31]
Naive Bayes Multinomial (NBM)	90.25%	[32]
PMI	66%	[33]
Dictionary Based lexicon	80%	[34]
Sentiment Lexicon with BPANN	95%	[35]

Table 2 provides a comprehensive overview of sentiment analysis studies. IMDB reviews were analyzed, achieving accuracies of 81.0% to 82.9% with different supervised learning approaches [30]. Focused on movie and car brand reviews achieving 87.2% accuracy [31]. Obtained 90.25% accuracy on blog data using Naive Bayes Multinomial [32]. Reached 66% accuracy on reviews with a lexicon-based approach [33]. Achieved 80% accuracy on data mining based on customer review [34]. Utilized neural networks for 95% accuracy for data mining on the Web [35]. Lastly, a survey [36] achieved 82.30% accuracy using sentiment lexicon and Linear SVM on CNET software and IMDB reviews. Sentiment analysis using text analytics is one of the greatest methods for conducting market research. These illuminating statistics can help brands set themselves out from the competition. Businesses are starting to adopt AI-powered sentiment analysis as a vital tool [40].

4. FRAMEWORK OVERVIEW

The framework name as "Fine Grained Sentiment Analysis for Code-Mixed Languages" (*FGSACML*) is defined in this section. By using this framework, we can get the sentiment polarities as Strongly Positive, Softly Positive, Positive, Strongly Negative, Softly Negative, Negative, and Neutral. With this model we can identify the sentiment at sentence level. Sentiment analysis can be defined as categorization of text into different polarities. This framework mainly divided into 5 sections. 1) Dataset Generation 2) POS Text Preparation 3) Sentiment Detection and Feature Extraction 4) Polarity Classifier and 5) Performance Evaluation. This framework applies to new features around Sentiment Analysis. The FGSACML can be used in many aspects like: User behaviors in social media, Customer review and requirement, branding in multi-language, etc. Because it is based on multi-language, if there is a review, comments or any polarities in the multi-language, this can be used to identify customer need, user behaviors, etc.

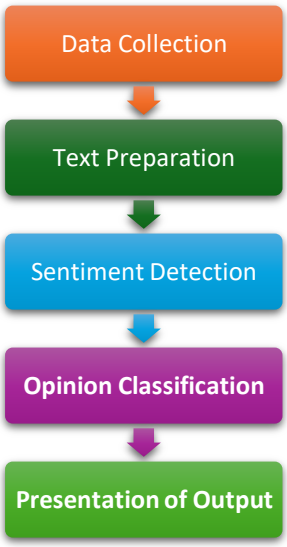

Fig.2. Framework Procedure

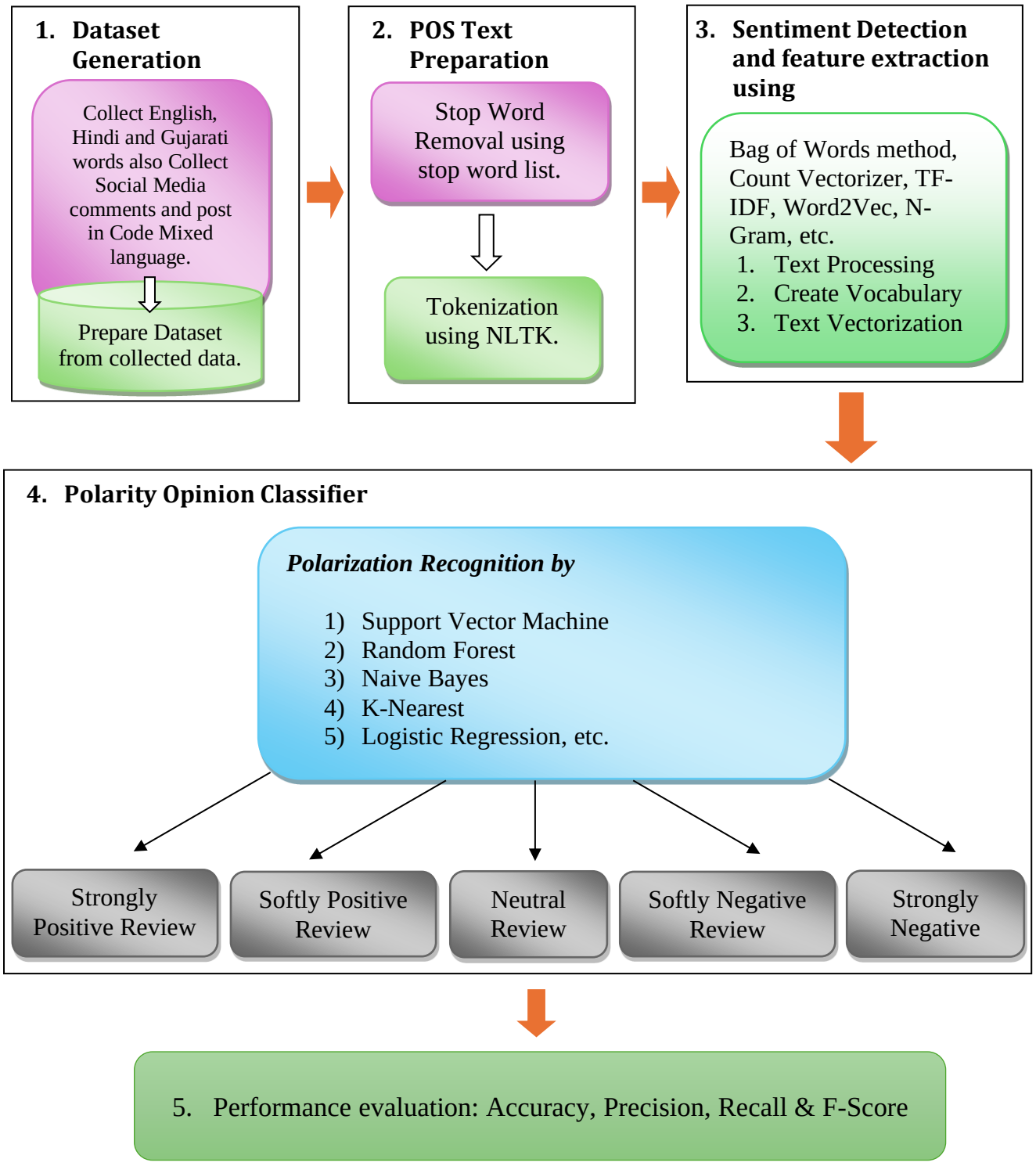

Fig.3. Framework for Fine Grained Sentiment Analysis for Code-Mixed Languages (FGSACML)

4.1 Dataset Generation

The Dataset is a collection of ordered data or information. An Information gathering will happen by study, analysis, observation, and measurement. The data can be generated from various E-Dictionary. Collect English, Hindi, and Gujarati words, also Collected Social Media comments and posts in the Code-Mixed language. In the dataset each row represents data, and every column corresponds to field (variable). The Dataset is structured. For this research we are Gathering post and comments from the social media like Twitter, Facebook, etc. and use it further in the next steps. From this step we are having 2 datasets as 1) Positive and Negative words and 2) list of Sentences with code mixed language.

4.2 POS Text Preparation

Long paragraphs, sometimes referred to as chunks of text, are broken up into tokens in this tokenization process, which are essentially sentences. You can further divide these statements into individual words. Take the statement "कल शाम को बाजार में I met to my teacher" as an illustration and separate tokenization comes: {“कल”, “शाम”, “को”, “बाजार”, “में”, “I”, “met”, “to”, “my”, “teacher”}

A basic goal in Natural Language Processing (NLP) is to assign a grammatical category (noun, verb, adjective, adverb, etc.) to every word in a document. This method enables computers to analyze and interpret human language more correctly by improving their grasp of phrase structure and semantics. Parts of Discourses Natural language processing (NLP) refers to the act of "tagging" each word in a text with a particular speech component (adverb, adjective, verb, etc.) or grammatical category. POS tagging is helpful in NLP applications for data extraction, machine translation, and named entity recognition, among other things.

Stop Word Removal is one of the preprocessing methods that is most usually used for different NLP applications. The idea is as straightforward as removing words that frequently occur in each document in the body. Stop Words are typically defined as articles and pronouns. These terms are not highly discriminative because they are irrelevant for some NLP tasks like classification and information retrieval. When indexing and retrieving entries based on a search query, search engines are programmed to reject terms that are designated as stop words. Examples of such words include "the," "a," "an," and "in", etc.

In this work, Tokenization and POS Tagging are used for text processing. Stop Word Removal is accomplished using NLP through the Python package known as NLTK.

4.3 Sentiment Detection and feature extraction

4.3.1 TF-IDF

The terminology frequency-inverse document frequency, or TF-IDF, is a frequently used technique for determining a word's relevance in a text. terminology frequency (t) is determined by dividing its total number of occurrences in a document by the document's word count. we can utilize IDF (Inverse Document Frequency) to determine a term's relevance. Certain phrases, such as "is," "an," "and" and so on, are used frequently but have no bearing on anything. Considering that N is the total number of documents and DF is the number of documents that contain the phrase t, the formula to compute IDF is IDF (t) = log(N/DF). TF-IDF is a more effective way to convert textual information representations into Vector Space Models (VSMs). The term frequency would be 10/250 = 0.04 if a text had 250 words and the word "Laptop" occurred 10 times in those 200 words. Similarly, suppose that out of 50,000 papers, only 500 mention the word "mouse." Then, IDF (Laptop) = TF-IDF (Laptop) will be equal to 50000/500 = 100 and 0.04*100 = 4.N-Grams

The textual characteristics for supervised machine learning algorithms will be formed by N-Gram. N-grams are consecutive strings of words or symbols. Technically, these can be identified as the successive group of elements in a document. They become important while handling text data in an NLP (Natural Language Processing) project [37]. Among all the theories in machine learning, the N-gram hypothesis is likely the most fundamental. A group of specific number of words is called an N-gram. There are 4 words in comment like "A Facebook comment post" is a 4-gram, "Facebook comment post" is a 3-gram (trigram), and "Facebook comment" is a 2-gram (bigram). But first, we need to look at the probabilities that are used in n-grams [38]. The value of the sentence's next word, p(w|h), can be predicted by an accurate N-gram model. A model called a unigram only considers the frequency with which a word occurs without considering any words that came before it. Bigram is a model that predicts the current word using only previous words. A trigram model that considers the two preceding words. After generalizing, the above equation can be calculated as:

$P\ (wn|w1:n-1) \approx\ P\ (wn|wn-N+1:n-1)$ [39]

4.4 Polarity Opinion Classifier

4.4.1 Support Vector Machine (SVM)

SVM is a machine learning supervised (feed-me) procedure that can be used to resolve regression or classification issues. Regression is the forecast of a continuous value, whereas classification is the forecast of a label or group. SVM is used for classification to locate the hyper-planes that differentiates the classes we plotted in n-dimensional space. The functioning for classification in support vector machines as

per the diagram. By using the opinion classifier, we can get the polarities of the input sentence.

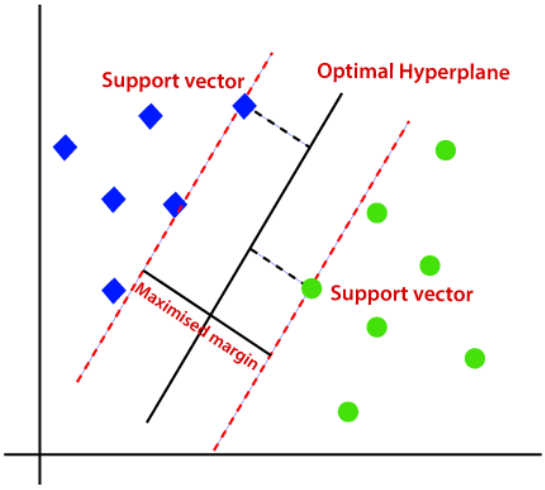

Fig.4. Support Vector Machine classification.

4.5 Performance Evaluation

This model can evaluate the model performance by accuracy, precision, recall and F-score, it is crucial to consider the trade-offs between these metrics and the problem's context. Below is an explanation of each:

4.5.1 Accuracy

Accuracy is defined as the proportion of correctly detected instances among all instances. Although it's an easy-to-understand statistic, it can be deceptive, particularly in datasets that are unbalanced and have a dominant class. For instance, even if a model with 95% accuracy that classifies everything as class A might not be particularly useful if 95% of your data falls into class A and only 5% into class B.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Negatives + False\ Positives}$$

4.5.2 Precision

The percentage of true positives, or positively anticipated cases, among all positively predicted cases is known as precision. Its main goal is to reduce false positives. When the cost of false positives is significant, precision matters. For instance, you wish to reduce the quantity of valid emails flagged as spam in a spam email detection system.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

4.5.3 Recall

Recall measures the proportion of true positives among all actual positives. Its main goal is to reduce misleading negative results. Recall becomes essential when the cost of false negatives is high. For instance, you wish to reduce the amount of real disease cases that are overlooked in a medical diagnosis system.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

4.5.4 F1-Score

The F-score calculates the harmonic mean of these two metrics, assigning equal weight to recall and precision. It offers a harmony between recall and precision. When you wish to consider both false positives and false negatives, the F1-score comes in handy. However, it might not be suitable for all scenarios, especially when you have imbalanced classes.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

5. CONCLUSIONS

This research study concludes with a thorough framework for fine-grained sentiment analysis on code-mixed language that is especially designed to comprehend the activities of social media users. There are several frameworks available for Sentiment Analysis but none of them worked on Fine Grained and with Code-Mixed language. Therefore, this framework (*FGSACML*) has a futuristic approach and has not been done before. The framework tackles the difficulties brought about by the dynamic and multilingual character of social media discourse by using a methodical approach that combines linguistic elements, machine learning techniques, and domain-specific knowledge. The model offers user attitudes, preferences, and interactions of sentiment analysis on code mixed language. This study advances our knowledge of code-mixed language sentiment analysis and applies new feature models for improving user engagement, sentiment tracking, and social media monitoring. This framework's continues development and improvement offers bright opportunities for improving our understanding of social media behavior in the digital era.

REFERENCES

[1] A. Dingli, L. Mercieca, R. Spina, and M. Galea, "Event detection using social sensors," in 2nd International Conference on Information and Communication Technologies for Disaster Management, 2015.

[2] S. Takeshi, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in 19th international conference on World wide web, ACM, 2010.

[3] J. Guerrero, J. Olivas, F. Romero, and E. Viedma, "Sentiment analysis: a review and comparative analysis of web services," Information Sciences, vol. 311, pp. 18–38, 2015.

[4] Lamba, Manika & Margam, Madhusudhan. (2022). Sentiment Analysis. 10.1007/978-3-030-85085-2_7.

[5] J. Guerrero, J. Olivas, F. Romero, and E. Viedma, "Sentiment analysis: a review and comparative analysis of web services," Information Sciences, vol. 311, pp. 18–38, 2015

[6] S. Wan and C. Paris, "Improving government services with social media feedback," in 19th international conference on Intelligent User Interfaces, pp. 27–36, 2014.

[7] K, K. P. (2020, August 9). A Literature review on application of sentiment analysis using Machine learning techniques. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3674982

[8] Yaakub, M. R., Latiffi, M. I. A., & Zaabar, L. S. (2019). A review on sentiment analysis techniques and applications. IOP Conference Series: Materials Science and Engineering, 551(1), 012070. https://doi.org/10.1088/1757-899x/551/1/012070

[9] Ligthart, A., Çatal, Ç., & TekiNerdoğan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review, 54(7), 4997–5053. https://doi.org/10.1007/s10462-021-09973-3

[10] Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. Applied Sciences, 13(7), 4550. https://doi.org/10.3390/app13074550

[11] Rao, A. C., Rao, A. C., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

[12] Ghosh, S., Hazra, A., & Raj, A. (2020). A comparative study of different classification techniques for sentiment analysis. International Journal of Synthetic Emotions, 11(1), 49– 57. https://doi.org/10.4018/ijse.20200101.oa

[13] Xu, Q., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. Decision Analytics Journal, 3, 100073. https://doi.org/10.1016/j.dajour.2022.100073

[14] Aspect based sentiment analysis on product reviews. (2018, December 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9096796

[15] Yan-Yan Z, Bing Q, Ting L (2010) Integrating intra-and inter-document evidences for improving sentence sentiment classification. Acta Autom Sinica 36(10):1417–1425

[16] Moreo A, Romero M, Castro J, Zurita JM (2012) Lexicon-based comments-oriented news sentiment analyzer system. Expert Syst Appl 39(10):9166–9180

[17] Joshi, Aditya, et al. ―A Fall-Back Strategy for Sentiment Analysis in Hindi: A ... - IIT Bombay.‖ www.Cse.Iitb.Ac.In, www.cse.iitb.ac.in/~adityaj/HindiSentiWordnet_AdityaJ.pdf.

[18] HOMS: Hindi opinion mining system. (2015, July 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/7232906

[19] Sentiment analysis of movie review data using Senti-lexicon algorithm. (2016). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/7912069

[20] Kour, K., Kour, J., & Singh, P. (2020). Lexicon-Based Sentiment analysis. In Lecture notes in electrical engineering (pp. 1421–1430). https://doi.org/10.1007/978-981-15-5341- 7_108

[21] Joshi, Vrunda, and Vipul Vekariya. An Approach to Sentiment Analysis on Gujarati Tweets. Vol. 10, no. 5, 2017, pp. 1487–1493.

[22] Ghosal, T., Das, S., & Bhattacharjee, S. (2015). Sentiment analysis on (Bengali horoscope) corpus. www.researchgate.net/publication. https://doi.org/10.1109/indicon.2015.7443551

[23] Kaur, A., & Gupta, V. (2014). N-Gram based approach for opinion mining of Punjabi text. In Lecture Notes in Computer Science (pp. 81–88). https://doi.org/10.1007/978-3-319- 13365-2_8

[24] Lexicon based sentiment analysis system for malayalam language. (2017, July 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8282571

[25] Se, S., Ravi, V., Kumar, M. A., & Soman, K. P. (2016). Predicting the Sentimental Reviews in Tamil Movie using Machine Learning Algorithms. Indian Journal of Science and Technology, 9(45). https://doi.org/10.17485/ijst/2016/v9i45/106482

[26] Polarity detection of Kannada documents. (2015, June 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/7154810

[27] Sentiment Analysis of Malayalam film review using machine learning techniques. (2015, August 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/7275974

[28] SentiMa - Sentiment extraction for Malayalam. (2014, September 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/6968548

[29] Patil, A., & Gupta, S. (2021). A review on sentiment analysis approaches. ResearchGate. https://www.researchgate.net/publication/349945234_A_Review_on_Sentiment_Analysis_Approaches

[30] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. arXiv (Cornell University). https://arxiv.org/pdf/cs/0205070v1

[31] Pang, B., & Lee, L. (2004). A Sentimental Education: sentiment analysis using subjectivity summarization based on minimum cuts. arXiv (Cornell University). https://arxiv.org/pdf/cs/0409058

[32] Boiy, Erik, Hens, & Marie-Francine. (2007). Automatic sentiment analysis in on-line text. ResearchGate. https://www.researchgate.net/publication/37676552_Automatic_Sentiment_Analysis_in_On-line_Text

[33] Peter, P., & Turney. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. Aclanthology, 417–424. https://research.microsoft.com/apps/pubs/default.aspx?id=65541

[34] Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. cs.uic.edu. https://doi.org/10.1145/1014052.1014073

[35] Sharma, A., & Dey, S. (2012). An artificial neural network based approach for sentiment analysis of opinionated text. ACMDL. https://doi.org/10.1145/2401603.2401611

[36] Tsytsarau, M., & Palpanas, T. (2011). Survey on mining subjective data on the web. Data Mining and Knowledge Discovery, 24(3), 478–514. https://doi.org/10.1007/s10618-011- 0238-6

[37] V, N. (2022) What are N-grams and how to implement them in python? Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2021/09/what-are-n-grams-and-how-to-implement-them-in-python/ (Accessed: February 25, 2023).

[38] Srinidhi, S. (2020) Understanding word N-grams and n-gram probability in natural language processing, Medium. Towards Data Science. Available at: https://towardsdatascience.com/understanding-word-n-grams-and-n-gram-probability-in-natural-language-processing-9d9eef0fa058 (Accessed: February 25, 2023).

[39] Jurafsky, Daniel, and James Martin. Speech and Language Processing. web.stanford.edu/~jurafsky/slp3/old_dec21/.

[40] Taherdoost, H.; Madanchian, M. Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research. *Computers* 2023, *12*, 37. https://doi.org/10.3390/computers12020037