



## DEEP LEARNING BASED SENTIMENT ANALYSIS OF CODE MIXED GUJARATI-ENGLISH TEXT—A REVIEW

**Mr. Chirag D. Shah**, Assistant Professor, KRP Darjee Indo-American Institute Of Computer Science, Vyara.

**Dr Shailesh A. Chaudhari**, Assistant Professor, JP Dawer Institute Of Information Science and Technology, VNSGU, Surat

### Abstract

NLP of people's views/opinions about a person/product/event/country expressed on various social media platforms is actively used by political parties, marketing companies etc. There has been a humongous escalation of Internet users which intend to use their native language while expressing their opinions or use combination of English and their native language. Indian constitution recognizes 22 languages which are used by various states/regions, people use their native language with combination of English to express their views. NLP of Gujarati language code mixed with English is meager, a few work have been done using machine learning techniques. NLP applications for sentiment analysis using deep learning have been greatly explored in English language as a source. So, it has become evidentially necessary to do sentiment analysis of code-mixed Gujarati-English text. This paper reviews current state of work in English and code mixed English text and need of deep learning based analysis of code mixed Gujarati-English text. The introduction section gives conception of Gujarati-English code mixed text using examples. The second section elaborates current techniques with their limitations and need of deep learning based architecture. The third section shows review of related work. The fourth section is of conclusion and future work that authors wish to carry in future.

**Keywords:** NLP, BERT, DNN, CNN, RNN, LSTM, Glove, TF-IDF, ML

### Introduction

With the growth in usage of social media in India from 50.44% in 2020 to estimated 67.40% in 2025, there has been increase in expression of views on different social media platforms such as Twitter, Facebook, Youtube, Whatsapp, and Telegram. According to Eighth schedule of constitution of India, there are 22 scheduled languages used by different states. Also English is used as an official language of communication between states and central government, parliament and also in judiciary system. When expressing their views about politics, games, weather, entertainment, education, tourism or any contemporary things people often uses intermixing of English and some regional language.

Gujarati is an Indo Aryan family language following Abugidas writing and is used by people of Gujarat and around 33% of Indian diaspora worldwide. The use of Gujarati language is over 700 years old and is currently used by around 60 million people in world. With increasing use of social media platforms, Gujarati-English code mixed/switched usage is commonly found.

While expressing their views users often get flexible with written expression, they use more than one language and often transliterates words of their domestic language into English. They convey their views by writing in multiple codes or mix up languages they are comfortable in. Code switching is use of more than one language in a single conversation; it has types such as inter-sentential and intra-sentential. Code mixing is a term which is closely related to code switching in which the change of language happens in a single utterance. Code switching happens between sentences (inter-sentential). whereas in code mixing, we insert some words/phrases of one language into the script of the other in a single sentence (intra-sentential).

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, appraisals, attitudes, and emotions toward entities and their attributes expressed in written text. The entities can be products, services, organizations, individuals, events, issues, or topics. Since, it mainly focuses on written text, it is an active field of research in NLP. Nowadays, opinions are central



to people activities ranging from buying a product to voting of elections. Thus, with growth of social media and proliferation of code mixing, the sentiment analysis is an active area of research.

### Code Mixed Text SA

Code mixed text applies grammar and vocabulary of two or more languages thus creating novel sentence structures. It refers to embedding of linguistic units such as words, phrase and morphemes of one language into utterance of another language. Transliteration of words in Gujarati like પહેલા to roman script can be like pahela, pehela, pahla, pehla. If we think of code mixing in Gujarati-English text it can be divided into 4 categories.

### Devanagari script(Gujarati) with transliteration of English words

Here, the sentence is mainly in Gujarati with few phrases or words of English inter mixed in conversation.

આજે પંજાબ કિંગ્સ અને કોલકાતા નાઇટ રાઇડર્સ વચ્ચે યોજાશે ડબલ હેડરની પહેલી મેચ

In above example, English words like કિંગ્સ, ડબલ ,હેડર, નાઇટ, રાઇડર્સ, મેચ are written in Gujarati.



Tv9 Gujarati @tv9gujarati · 1h

Tender Today : બોટાદ નગરપાલિકામાં બોક્સ કલવર્ટ બનાવવાના કામ માટે લાખો રૂપિયાનું ટેન્ડર જાહેર

#Botad #Boxculvert #Tendertoday

In above tweet(credit: Tv9 Gujarati), English words like બોક્સ, કલવર્ટ , ટેન્ડર are written in Gujarati.

### Gujarati transliterated in Roman script

Here, the sentence is mainly in Gujarati with transliteration in English and few phrases or words of English inter mixed in conversation.

\* Story of my life \* - Bau bhukh laagi che lya! - Bau khavaai gayu lya! – Repeat

Here, phrase story of my life and word repeat are from English language.



@ahmadmanasiya7693 5 months ago

Tamne jovu chu paras bhai hu Amari school ma tamari motivation Apu 🍌🍌🍌🍌🍌❤️❤️❤️

2 Reply

In above Youtube comment(credit:ahmadmanasiya7693), English words like school, motivation are written in transliterated statement.

### Devanagari script(Gujarati) script with English words

Here, the sentence is mainly in Gujarati written in Devanagari and few phrases or words of English inter mixed in conversation.

Good Luck મેળવવા માગતા લોકો એ અરીસાને ખોટી દિશામાં ગોઠવવાની ભૂલ ન કરવી જોઈએ, વાંચો શું કહે છે વાસ્તુનો સાચો નિયમ

Here, English language phrase Good Luck is code mixed in Gujarati sentence.



@parmarpravinkumar4950 1 year ago

હું આજ સુધી ઘણા સારા મોટીવેશન સ્પીકર સાંભળ્યા છે પણ સૌથી વધુ રસપ્રદ વાતો આ વિડીયોમાં છે... Best Speak Safin Sir

112 Reply

In above Youtube comment(credit:parmarpravinkumar4950), English phrase Best Speak Safin Sir is code mixed in Gujarati sentence.



### **Roman script (English) with Gujarati words**

Here, the sentence is primarily in English with some Gujarati phrases inter mixed in the sentence.

I am proud of you Khajurbhai, બૂબ સરસ

Here ,Gujarati phrase બૂબ સરસ is intermixed in the English sentence.

### **Existing Techniques of Sentiment Analysis**

The main goal of SA is to automatically identify whether a given text has a positive, negative or a neutral impression. Existing techniques with their limitations are as under.

- Lexicon based techniques were originally used, we have a dictionary of words with polarities as positive,negative,neutral and sentiment scores. We calculate sentiment scores of each word in a sentence by assigning them sentiment labels, then based on this counting we calculate overall sentiment. We can develop lexicon manually or use predefined dictionary of terms such as TextBlob, AFINN, VADER, LIWC, ANEW, SentiWordNet and WordNet. They are easily accessible and so not require implementation using advanced algorithms. They can be either dictionary based(collection of positive, negative words) or corpus based(uses statistical techniques).

#### **Limitation of lexicon approach**

(a) Different languages may have different meaning of same pronounced words. Eg. “man” in English means a male adult whereas the same word mean “mind” in Gujarati. Lexicon based models cannot identify words based on language, this creates problem in code mixed text.

(b)Sentiment may vary depending on the context in which given word is used Eg. “bowl” has different meaning in Cricket and meaning, lexicon based models cannot capture the context.

(c) Sarcasm detection is also difficult in lexicon based approach.Eg.”Rohit Sharma goes for golden duck again, Great!”. This is statement is sarcastic, but will be classified as positive due to word Great.

(d)These models also do not handle slangs, creative spellings and negation very well. So, lexicon based models are not ideal for social media text.

(e) They also require manual labelling of the text and requires domain knowledge for labelling.

- Machine learning based techniques do not rely on manually crafted rules. They allow computers to learn new task without the need to explicit programmed. They can be trained to understand things like context, sarcasm, misapplied words etc. They can adapt to a change in input and can be trained to understand context, sarcasm, creative spellings and slangs. There are supervised (LR,SVM,RF,NB,DT etc.), unsupervised(K-means,Apriori) and semi supervised techniques based on labeling of data. We have to do go through text preprocessing, feature engineering, model selection, training and evaluation phases.

#### **Limitation of ML approach**

(a) The selection of features is done manually and accuracy depends on how features are chosen. Also they do not perform well when volume of data is big.

(b)They find it difficult to disambiguate text with mixed sentiments. For eg. “Rohit is a good batsman, but he fails in crucial matches”, statement has mixed filling which is difficult to capture in ML models.

(c) While doing annotation, if training data is biased predictions goes wrong. Also, they suffer from under fitting (when dataset is small) and over fitting (when training is extensive and model starts learning from noise) problems.

(d)These models cannot capture long range dependencies between words.

#### **The emergence of Deep Learning**

Deep learning is an emerging field that automatically learn features by application of artificial neural networks to learning tasks using network of multiple layers. It is inspired by structure of biological brain which has neurons working as information processing units. The network learns by adjusting weights of neurons between layers, the features are learned automatically without the need of human intervention. So, the feature engineering is not hand crafted. They are capable of learning on a new



task by applying small changes to an existing model. They can be either feedforward or recurrent based on connection between output and input layers. It consists of at least three layers (input, hidden and output). In deep learning, models such as DNN, CNN, RNN (LSTM, BiLSTM, and GRU) and transfer learning are used. Deep learning generates dense embedding as compared to sparse matrices generated by traditional methods.

### Literature

The authors did one of the first attempts to classify Gujarati text using Twitter dataset. They used SVM to classify inputs into positive and negative polarity. The authors have performed sentiment analysis of 40 Gujarati tweets using POS tagging. They used SVM algorithm for classifying tweets as positive or negative and achieved accuracy of 92% on this small dataset. (Joshi, n.d.)

The authors performed SA of Gujarati film reviews using machine learning techniques. They have done sentiment analysis of movie review collected from data collected from a Gujarati movie review website. They used machine learning approach with TF-IDF and Count Vectorizer. They have used MNB and K-nearest neighbours to do review classification. They achieved accuracy of 87.14 with MNB and 81.43 with KNN using TF-IDF whereas 81.43 with MNB and 81.43 with KNN using CV. They also found that, there is no standard dataset available for Gujarati film reviews, so they prepared dataset of 500 Gujarati movie reviews. (Shah et al., 2022)

The authors have performed Gujarati codemixed language discovery which is written in a sentence either in a transliterated or original forms mixed with Hindi and English. They have scrapped Youtube comments for their research work. They have used Logistic Regression, K-nearest Neighbor, Naive Bayes, Decision Tree, Support Vector Machine, and Random Forest for language identification task. The SVC gave highest accuracy of 92%. (Kazi et al., 2020)

The authors notice scarcity of resources available for Gujarati language and have created a corpus for sentiment analysis of Gujarati text extracted from Twitter. The text annotation was done by native language speakers. They have created an openly accessible dataset GSAC using 6575 tweets. They have tested this dataset on various machine learning and deep learning models and found that GujaratiBERT and IndicBERT has best accuracy of 0.64 and 0.66 respectively. (Gokani & Mamidi, 2023)

The authors have noted that Gujarati is under resourced language for sentiment analysis task. They have developed a corpus called GujaratiSentiWordNet on the basis of HindiSentiWordNet and IndoWordNet. They have achieved accuracy of 52.72% and 52.95% using unigram presence method and simple scoring method respectively. (Computer Science and Engineering Department, Institute of Technology, Nirma University, India Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA), CHARUSAT, Changa, India et al., 2019)

The authors have performed sentiment analysis on GujSentiWordNet using lexicon based and machine learning based approaches. In machine learning approaches they have used logistic regression (LR), random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), naive Bayes (NB) classifiers with TF-IDF and Count Vectorizer. They found that machine learning models performed better than lexicon based approaches with minimum 75% accuracy except KNN. (Department of Computer Engineering, Sarva Vidyalaya Kelavani Mandal managed Vidush Somany Institute of Technology and Research, Kadi, India et al., 2023)

The authors performed SA of Hindi-English and Bengali-English code mixed text. They used Twitter4J API to collect code mixed data. Out of all teams, the best performing team obtained maximum macro average f-score of 0.569 and 0.526 for HI-EN and BN-EN datasets, respectively. The team used word and character level n-grams as features and SVM for sentiment classification. (Patra et al., 2018).

The authors performed sentiment analysis on Dravidian family of languages. They have scrapped data from "Dravidian-CodeMix-FIRE 2021" Youtube comments. They initially developed a hybrid model using CNN and Bi-LSTM that outperforms machine learning models by a little margin. To get more





accuracy, they developed a deep learning model based on BERT. The two BERT models achieved F1 score of 0.64(Tenserflow) and 0.66(Ktrain) respectively as compared to 0.55 of hybrid models and 0.50 of SVM(Dutta et al., n.d.).

The authors collected Gujarati tweets having 8 emotions and created a dataset. They classified tweets according to machine learning and hybrid techniques. They found that hybrid techniques outperform machine learning techniques like SVM for feature generation(Sevak & Bhadoria, 2021).

The authors used machine learning model and deep learning model to do sentiment analysis of Hindi-English code mixed text. They combined output of both the models to predict final sentiment of the statement. They had dataset of 3979 sentences and achieved accuracy of 84.9 percent(Tanusha et al., 2020).

The authors have performed sentiment analysis on code mixed Tamil-English text. They have classified text as s positive, negative, neutral, mixed-feeling and non-Tamil for Tamil and Malayalam languages. They have used 15,744 Youtube comments as their dataset. In their approach, a deep learning based Bi-LSTM model has been used for the classification purpose. They have achieved precision of 0.53% precision on validation data and 0.33% on test data(*SA-SVG@Dravidian-CodeMix-FIRE2020: Deep Learning Based Sentiment Analysis in Code-Mixed Tamil-English Text*, n.d.).

The authors have performed offensive language identification and sentiment analysis on Malayalam-English cide mixed dataset using deep learning algorithms. The code-mixed data set for SA is classified into 5 classes: Positive, Negative, Unknown state, Mixed feelings, and not-Malayalam. The shared task for SA is constituted of 4851 code-mixed social media comments/posts in the training set, 540 comments in the validation data set and 1348 comments in the test data set. They have applied simple deep learning models, bidirectional models, hybrid models, and transformers. They evaluated performance based on F1-score(Thara & Poornachandran, 2022).

The authors performed SA on Facebook pages with 3879 comments. Their deep learning architecture consists of two major parts. The first part learns the sub-word level features from input character sequences. The other part used trained word embedding as input and learns the word level features. They have achieved accuracy of 69.845%(Mukherjee, 2019)

The authors collected data from social media sites like Twitter, Facebook, Youtube, bookmyshow,newspaper etc. They converted the preprocessed code mixed text in to English using Google API converter. Then, they used deep learning method to classify the sentiment(Pawar et al., 2021).

The authors used dataset of 2,50,000 tweets from which they removed Devanagari and pure English tweets and considered only Hinglish tweets. They classified them into 6 emotions (Happy, Sad, Anger, Fear, Disgust,and Surprise). They used FastText word embedding. They tested different deep learning algorithms and found that Bi-LSTM (Attention) performed best with accuracy of 87.78%(Das & Singh, 2023).

The authors have performed sentiment analysis of code mixed Tamil-English text using Bi-directional LSTM on FIRE-2021 dataset. They achieved weighted average precision, recall and F1-score values are 0.54, 0.57, and 0.55, respectively on the validation dataset. However, on the test dataset, the weighted precision, recall and F1-score values are 0.544, 0.566, and 0.552, respectively(Roy & Kumar, n.d.).

The authors have developed recommender system for streaming services. They used MARD and Amazon review dataset. They have used pre-trained BERT to create feature vectors. The reviews and genre data are fed into the BERT model to generate the feature vectors. In the case of genres, the vectors are used to compute the weight of the user similarity, while feature vector obtained from reviews are the input to the hybrid deep learning models that perform the sentiment classification. They have used two hybrid deep learning models CNN and LSTM for their prediction(Dang et al., 2021).

The authors performed sentiment analysis of movie reviews. They compared performances of various machine algorithms like RF, NB, SVM, KNN, CNN and hybrid CNN+SVM. They found that hybrid



CNN+SVM model achieved highest accuracy of 92%(Department of BCA, Gurunanak College, Chennai, India et al., 2019).

The authors have performed deep learning based sentiment analysis on IMDB dataset of 50,000 movie reviews. They have applied CNN, RNN and LSTM algorithms on the given data. CNN gave the accuracy of 88.22%, while RNN and LSTM have accuracy of 68.64% and 85.32% respectively(Cen et al., 2020).

The authors have performed deep learning based sentiment analysis and offensive language detection on code mixed Tamil-English data. The data set consists of a training set which consists of 32,604 tweets, validation set consists of 4076 tweets and test set consists of 4076 tweets. They have used Glove and BERT for word embedding. They have used pre-trained models like Logistic regression, CNN, BERT, RoBERTa, Bi-LSTM and Adapter-Bert for text classification. Among the obtained results Adapter BERT performs better than other models with the accuracy of 65% for sentiment analysis and 79% for offensive language identification(Shanmugavadivel et al., 2022).

The authors have performed SA of Hindi movie reviews collected from websites and newspapers. The dataset was manually annotated. They used CNN to achieve accuracy of 95% and they also compared model performance with traditional machine learning based models(Rani & Kumar, 2019).

The authors have used combination of machine learning model which is MNB and end-to-end character ngram based deep learning model which uses LSTM to classify sentiments of Hindi-English code mixed text into three polarities. To predict final sentiment, they have combined output of both the models. (Jhanwar & Das, 2018).

The authors have performed sentiment analysis with language identification on low resource Hindi-English text using BERT based models. They used datasets like Icon Sentiment Dataset, Sentiment Dataset, Hate Speech Dataset, Emotions Dataset and HASOC Dataset. They found that Hing based BERT model outperformed vanilla BERT models.(Takawane et al., n.d.)

The authors have used Hinglish dataset of tweets. They first used Baseline Monolingual Systems, then they used Supervised Classification with cross lingual embeddings and at last they used Transfer learning based models by applying trained dataset of English on code mixed Hinglish. For supervised architectures, MUSE SeedDict gave an impressive precision of 0.66 whereas for transfer learning VecMap SeedDict gave an precision of 0.56.(Singh & Lefever, n.d.)

The author has performed sentiment analysis on code mixed Kannada and Malayalam languages using transformer based models. For Kannada code mixed text transformer based models achieved a weighted average F1 score of 0.66 whereas for Malayalam code mixed text they used ensemble deep learning mode with weighted F1 score of 0.72.(Roy, 2023)

They have noticed that machine learning models like SVN, NB suffers from performance issues while having large data. So, they applied LSTM for sentiment analysis of E-Commerce comments on Chinese dataset.(Qiwei & Xiaochao, n.d.)

The authors have experimented for Malayalam-English and Tamil-English code mixed text, with models like Naïve Bayes, Decision tree, Random Forest, AdaBoost and Long Short Term Memory (LSTM) for sentiment classification. The models trained for hate speech and offensive content identification are Naïve Bayes, SVM, Decision tree, Random Forest, Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). They also developed an ensemble of Machine Learning classifiers for both challenges. They found that LSTM gives higher accuracy of 88% for Dravidian code mixed text.(Swaminathan et al., n.d.)

The authors have used M-BERT and XLM-RoBERTa to do sentiment analysis of Dravidian code mixed data. They found that deep learning model M-BERT gives 0.73 F1 score for Malayalam and 0.64 for Tamil.(Huang & Bai, n.d.)

The authors have proposed a Bi-LSTM for sentiment analysis for Kannada-English (Kn-En), Malayalam-English (Ma-En), and Tamil-English (Ta-En) code-mixed texts. They have achieved F1



score of 0.563 for Ma-En, 0.604 for Ta-En and 0.369 for Kn-En on test dataset. (Anusha & Shashirekha, n.d.)

The authors have performed sentiment analysis of Persian-English code mixed text. They have noted that Naïve Bayes, Random Forest and Bi-LSTM achieves accuracy of 60.88, 62.12 and 66.17 respectively. The deep learning based Bi-LSTM outperforms other models. (Sabri et al., 2021)

The authors have performed sentiment analysis on code mixed Tamil, Malayalam, and Kannada comments of Youtube. They have used Deep Neural Network for sentiment analysis. They have achieved

F1 score of 0.42 for Tamil, 0.52 for Malayalam and 0.48 for Kannada. (Fadil & S K, n.d.)

The authors have marked limitation of machine learning models in feature extraction and also the need of cross language sentiment classification. (Raghunathan & Saravanakumar, 2023)

## Conclusion

From literature review it is clearly evident that there are a few large datasets available for Gujarati-English code mixed text and that research in code mixed Gujarati-English text is only at inception stage. The authors of this paper intend to develop a deep learning based model for sentiment analysis of code mixed Gujarati-English text using self-developed dataset. The first step of data collection will be done by scrape of Youtube, Twitter, Facebook comments and tweets. In second step, we will preprocess and tokenize data at word level. Next step would be to generate word embedding for Gujarati and English languages. In next step, we will create a deep learning architecture using RNN with LSTM. The last step would be to train/test the model and evaluate performance using different metrics.

## References

- [1] Anusha, M. D., & Shashirekha, H. L. (n.d.). *BiLSTM-Sentiments Analysis in Code-Mixed Dravidian Languages*.
- [2] Cen, P., Zhang, K., & Zheng, D. (2020). Sentiment Analysis Using Deep Learning Approach. *Journal on Artificial Intelligence*, 2(1), 17–27. <https://doi.org/10.32604/jai.2020.010132>
- [3] Computer Science and Engineering Department, Institute of Technology, Nirma University, India Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA), CHARUSAT, Changa, India, Gohil, L., Patel, D., & Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA), CHARUSAT, Changa, India. (2019). A Sentiment Analysis of Gujarati Text using Gujarati Senti word Net. *International Journal of Innovative Technology and Exploring Engineering*, 8(9), 2290–2292. <https://doi.org/10.35940/ijitee.I8443.078919>
- [4] Dang, C. N., Moreno-García, M. N., & De La Prieta, F. (2021). Using Hybrid Deep Learning Models of Sentiment Analysis and Item Genres in Recommender Systems for Streaming Services. *Electronics*, 10(20), 2459. <https://doi.org/10.3390/electronics10202459>
- [5] Das, S., & Singh, T. (2023). Sentiment Recognition of Hinglish Code Mixed Data using Deep Learning Models based Approach. *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 265–269. <https://doi.org/10.1109/Confluence56041.2023.10048879>
- [6] Department of BCA, Gurunanak College, Chennai, India, K\*, Mrs. Raviya., Vennila.S, Dr. M., & Department of BCA, Gurunanak College, Chennai, India. (2019). A Novel Deep Learning Based Sentiment Analysis of Movie using Hybrid CNN\_SVM Algorithm. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(4), 12391–12394. <https://doi.org/10.35940/ijrte.D7115.118419>
- [7] Department of Computer Engineering, Sarva Vidyalaya Kelavani Mandal managed Vidush Somany Institute of Technology and Research, Kadi, India, Shah, P., Swaminarayan, P., Faculty of Information Technology and Computer Science, Parul University, Vadodara, India, Patel, M., &



- Department of Computer Engineering, Gandhinagar University, India. (2023). Sentiment Classification for Film Reviews in Gujarati Text Using Machine Learning and Sentiment Lexicons. *Journal of ICT Research and Applications*, 17(1), 1–16. <https://doi.org/10.5614/itbj.ict.res.appl.2023.17.1.1>
- [8] Dutta, S., Agrawal, H., & Roy, P. K. (n.d.). *Sentiment Analysis on Multilingual Code-Mixed Kannada Language*.
- [9] Fadil, N. M., & S K, L. (n.d.). *Sentiment Analysis of YouTube comments in Dravidian Code Mixed Language using Deep Neural Network*.
- [10] Gokani, M., & Mamidi, R. (2023). GSAC: A Gujarati Sentiment Analysis Corpus from Twitter. *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, 129–137. <https://doi.org/10.18653/v1/2023.wassa-1.12>
- [11] Huang, B., & Bai, Y. (n.d.). *Sentiment Analysis on Multilingual Code Mixing Text with M-BERT and XLM-RoBERTa*.
- [12] Jhanwar, M. G., & Das, A. (2018). *An Ensemble Model for Sentiment Analysis of Hindi-English Code-Mixed Data* (arXiv:1806.04450). arXiv. <http://arxiv.org/abs/1806.04450>
- [13] Joshi, V. C. (n.d.). *An Approach to Sentiment Analysis on Gujarati Tweets*.
- [14] Kazi, M., Mehta, H., & Bharti, S. (2020). Sentence Level Language Identification in Gujarati-Hindi Code-Mixed Scripts. *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (ISSSC)*, 1–6. <https://doi.org/10.1109/iSSSC50941.2020.9358837>
- [15] Mukherjee, S. (2019). Deep Learning Technique for Sentiment Analysis of Hindi-English Code-Mixed Text using Late Fusion of Character and Word Features. *2019 IEEE 16th India Council International Conference (INDICON)*, 1–4. <https://doi.org/10.1109/INDICON47234.2019.9028928>
- [16] Patra, B. G., Das, D., & Das, A. (2018). *Sentiment Analysis of Code-Mixed Indian Languages: An Overview of SAIL Code-Mixed Shared Task @ICON-2017* (arXiv:1803.06745). arXiv. <http://arxiv.org/abs/1803.06745>
- [17] Pawar, R., Salunkhe, S., Dhongdi, S., Waghmare, V., & Rane, T. A. (2021). *Sentiment Analysis of Code Mixed Text*. 08(12).
- [18] Qiwei, C., & Xiaochao, L. (n.d.). Research on Text Sentiment Analysis of E-Commerce Comments Based on Deep Learning. *International Journal of Social Science and Education Research*.
- [19] Raghunathan, N., & Saravanakumar, K. (2023). Challenges and Issues in Sentiment Analysis: A Comprehensive Survey. *IEEE Access*, 11, 69626–69642. <https://doi.org/10.1109/ACCESS.2023.3293041>
- [20] Rani, S., & Kumar, P. (2019). Deep Learning Based Sentiment Analysis Using Convolution Neural Network. *Arabian Journal for Science and Engineering*, 44(4), 3305–3314. <https://doi.org/10.1007/s13369-018-3500-z>
- [21] Roy, P. K. (2023). A Deep Ensemble Network for Sentiment Analysis in Bi-Lingual Low-Resource Languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 3600229. <https://doi.org/10.1145/3600229>
- [22] Roy, P. K., & Kumar, A. (n.d.). *Sentiment Analysis on Tamil Code-Mixed Text using Bi-LSTM*.
- [23] Sabri, N., Edalat, A., & Bahrak, B. (2021). *Sentiment Analysis of Persian-English Code-mixed Texts* (arXiv:2102.12700). arXiv. <http://arxiv.org/abs/2102.12700>
- [24] *SA-SVG@Dravidian-CodeMix-FIRE2020: Deep Learning Based Sentiment Analysis in Code-mixed Tamil-English Text*. (n.d.).
- [25] Sevak, T., & Bhadoria, S. S. (2021). *Embedding And Classification Strategy For Sentiment Analysis On Mixed-Indic Social Media Text*. 18(5).





- [26] Shah, P., Swaminarayan, P., & Patel, M. (2022). Sentiment analysis on film review in Gujarati language using machine learning. *International Journal of Electrical and Computer Engineering (IJECE)*, 12(1), 1030. <https://doi.org/10.11591/ijece.v12i1.pp1030-1039>
- [27] Shanmugavadivel, K., Sathishkumar, V. E., Raja, S., Lingaiah, T. B., Neelakandan, S., & Subramanian, M. (2022). Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. *Scientific Reports*, 12(1), 21557. <https://doi.org/10.1038/s41598-022-26092-3>
- [28] Singh, P., & Lefever, E. (n.d.). *Sentiment Analysis for Hinglish Code-mixed Tweets by means of Cross-lingual Word Embeddings*.
- [29] Swaminathan, S., Ganesan, H. K., & Pandiyarajan, R. (n.d.). *Sentiment Analysis and Hate Speech Identification using Machine Learning, Deep Learning and Ensemble Models*.
- [30] Takawane, G., Phaltankara, A., Patwardhan, V., Patil, A., Joshi, R., & Takalikar, M. (n.d.). *Leveraging Language Identification to Enhance Code-Mixed Text Classification*.
- [31] Tanusha, P. V., Kayani, S., Sruthi, P., VishnuPriya, A., & Sukanya, V. (2020). *MULTILINGUAL TEXT CLASSIFICATION USING SENTIMENT ANALYSIS*. 8(5).
- [32] Thara, S., & Poornachandran, P. (2022). Social media text analytics of Malayalam–English code-mixed using deep learning. *Journal of Big Data*, 9(1), 45. <https://doi.org/10.1186/s40537-022-00594-3>