

Review on Sentiment Analysis of Indian Languages with a Special Focus on Code Mixed Indian Languages

Gazi Imtiaz Ahmad
Ph.D Scholar
Lovely Professional University
Phagwara, Punjab, India
gaziimtiaz@gmail.com

Dr. Jimmy Singla
School of CSE
Lovely Professional University
Phagwara, Punjab, India
jimmy.21733@lpu.co.in

Nikita
Former Assistant Professor
School of CSE, LPU
Phagwara, Punjab, India
nikita.jindal17@gmail.com

Abstract—Given the wide applicability of social media platforms, individuals increasingly turn to the web to seek and share information, opinions, comments and suggestions which results in proliferation of user generated large volume of text data available for interpretation. A large number of users in India write their feelings or emotions in more than one language, thereby a large volume of text data is made available for Natural Language Processing (NLP) researchers. Sentiment Analysis (SA) of code-mixed text provides useful information in the field of politics, marketing, business, health, sports and what not. During the past decade the work on Sentiment Analysis of Indian language textual data, particularly in Hindi has got momentum in contrast to code-mixed Indian language text. However, due to non-availability of language and vocabulary (linguistic and lexical) tools and annotated resources, the task of Sentiment Analysis of Indian Languages becomes somehow difficult. In this study an attempt has been made to provide a detailed summary of Sentiment Analysis of Indian languages with a special focus on code mixed Indian Languages.

Keywords—Natural Language Processing, Sentiment Analysis, code-mixed Indian languages, Machine Learning, Deep Learning

I. INTRODUCTION

With the increase in social media text, the researchers took interest in analyzing the data generated from social media platforms to infer desired information from it. Since India is a multilingual country, so is the text available on social media sites in different Indian languages. To provide better results in the field of marketing and business, people associated with this trade rely on the beliefs/views of other people about their products and services. On Social media sites such as Twitter and Face-book, people post their opinions, comments, suggestions in a free form, resulting in large volume of text data available for interpretation. People read through a paragraph/document written in a language which they know and can easily infer whether the text contains positive, negative or neutral feelings. However, for a computer which does not have the concept of spoken languages around the globe, it cannot interpret the text. Sentiment Analysis (SA) solves this problem of computers by using Natural Language Processing to recognize words in a document and to classify feelings/ emotions accordingly. Sentiment Analysis (opinion/ review mining) is the task of identifying, mining and classifying feelings, views, attitudes, sentiments, reviews expressed by people on different subjects [1],[33].The task of sentiment analysis include identification of objectivity and subjectivity of a textual input. If the text does not contain any opinion, it is objective

in nature whereas, a text containing feelings/emotions about a particular entity, it is specified as subjective [2]. e.g. “*It is raining*” conveys general information and is objective in nature whereas “*I am happy, it is raining*” conveys positive feeling about a particular event and hence subjective. Subjective sentences convey positive, negative or neutral feelings. In India large number of people are using social media sites for posting their opinions. According to a study of KPMG Group in India and Google in April 2017, Indian language users can grow up to 536 million by 2021. It has also been observed that they express their views on varying subjects ranging from politics to movies and to sports These opinions are often expressed in their own languages such as Hindi, Bengali, Tamil, Punjabi etc., as all the social media sites provide interface to a user to express a view in his/her own language. Also a large number of people express their feelings in mixed languages such as English-Hindi, English-Bengali, English-Tamil, and English-Urdu. So a large volume of data is available either in a single Indian language or in a code-mixed/code-switched language depending how they use two languages. This multilingual data is growing exponentially on the Internet, however due to unavailability of annotated resources, linguistic and lexical tools in the Indian languages, it becomes difficult for researchers to interpret the data through traditional sentiment analysis techniques.

This study is based on the relevant research studies emerged during last 10 years. The information was collected from Google Scholar, Science Direct, ACM Digital Library, IEEE Explore and Research-gate with least possibility of redundancy. Only papers of last 10 years of both quantitative and qualitative in nature were selected. Based on the search string on the above mentioned repositories, a total of 105 papers were retrieved which were reduced to 30 for final selection. The study shows how much work has been done in a particular Indian language for which Sentiment Analysis technique has been used. Also, how many resources in terms of linguistic, lexical and datasets are available for different Indian languages, including the work done using Sentiment Analysis for code-mixed Indian languages.

II. BACKGROUND

Sentiment Analysis identifies and extracts subjective information in a source material to help a person, entity or business to understand the social opinions of their product or service. The text data available on Internet in the form of emails, chats, social media, surveys, articles and documents is largely unstructured. Manual analysis of this data is usually difficult, time-consuming and expensive. Sentiment analysis system provides ways of analyzing this unstructured

data in an efficient manner, thereby saving a lot of time, labour and money. Owing to its wide area of applications, Sentiment analysis has become a key area of interest and research for business organizations. Sentiment Analysis is used for analyzing trends, evaluation of public opinions, identifying ideological bias, targeting advertisements, analyzing reviews of a product and services, opinions and reactions to ideas. Thus in a world where trillion bytes of textual data is generated on daily basis, sentiment analysis has become a key tool for making sense out of this data.

During the last 10 years, the users on the social media and microblogging sites has increased manifold. The data/text available on these sites in the form of feelings, emotions, reviews, opinions, suggestions, arguments, comments etc. attract Natural Language Processing researchers to deduce relevant information thereby making Sentiment Analysis one of the active research area in the field of NLP.

As the data in earlier stage was available in English language, therefore most of the work for creating resources, linguistic and lexicon tools and presence of data sets makes English language as the primary choice for SA. English being the dominant language around the world, therefore the work in the field of Sentiment Analysis is predominantly in English. However, as the Internet grows and people around the world start using Internet and Social Media, the data in other languages also start growing. Therefore, significant work has also been carried out in other non-English languages such as Arabic, Spanish, French, and Chinese etc. A large volume of datasets, linguistic and lexical resources are also available for these languages.

India is a large country with population of more than 1 billion and around half of the population is using Internet due to the availability of the same at cheaper rates. This gives rise to the data in Indian languages on the web at an exponential rate. Indian languages have two broader families viz. Indo Aryan family, Dravidian family. Almost 98% of languages spoken in India come from these two families [2]. Also a large number of population in India knows more than one language so the text available is either in a single Indian language or in code-mixed Indian languages such as English-Hindi. However, due to the absence of large volume of datasets, linguistic and lexical resources for Indian languages, very less contribution exists [3]. In 2010 [4] makes the first attempt in the field of NLP and SA for Indian languages. It was a first attempt of its kind for SA of Hindi language which opens way for interested persons to conduct SA on other Indian languages. The way of expression in Indian languages is quite different from the English language, as the sentence in English language starts with a subject followed by verb followed by object in contrast to the Indian languages sentences where the object precedes the verb. Moreover, Parts of Speech (POS) of a word gets changed when sentence is converted from English to Indian language. Therefore the Sentiment Analysis of Indian languages is different and difficult than the English language.

With the use of advanced Artificial Intelligence Techniques and advancements in deep learning algorithms, there has been considerable improvement in the field of sentiment analysis of textual data.

III. OVERVIEW OF SENTIMENT ANALYSIS

Sentiment Analysis of text involves a number of stages

starting from collection of data, preprocessing, detection and classification of sentiment, to the presentation of results. Sentiment Analysis can be applied at feature, sentence, and phrase or document level. In feature or aspect based sentiment analysis, we detect the features from the textual input and detect the sentiments expressed by people for each feature [5]. Sentiment Analysis at the sentence level involves two phases: firstly checking the objectivity and subjectivity of the sentence and then classifying the emotions/sentiments into positive, negative and neutral. Phrase level involves analyzing phrases in a sentence in deeper sense to classify them as positive, negative, neutral and objective. For document level Sentiment Analysis, it is necessary that the text contains sentiments about a single entity.

Three broad categories of Sentiment Analysis techniques exists which are Lexicon based, Machine Learning (ML) and Deep Learning. To test a particular approach, the researchers use a number of measures. These measures include Accuracy, Precision, Recall and F-measure. These four metrics are defined in terms of True Positive, True Negative, False Positive and False Negative.

A. *Lexicon based Sentiment Analysis*

The Lexicon based approach is based on some predefined rules to determine the sentiment from the text. In this approach, positive and negative feelings are counted to assign positive, negative or neutral score. This approach is used at sentence and document level. To construct a sentiment lexicon, manual construction, dictionary based or corpus based method is followed. The annotated resource consists of words and their corresponding sentiment polarity [6],[7]. The lexicon and dictionary based approach has paved way to machine learning approaches.

B. *Machine learning based Sentiment Analysis*

Machine Learning approach uses syntactic and semantic/pattern mining to identify sentiments. Machine Learning approaches are supervised, unsupervised and semi-supervised. Algorithms such as Naïve Bayes, Support Vector Machine, Decision Tree and k-nearest Neighbor are used in Supervised Learning. A training dataset is used by the classifier to learn and train itself to identify the features in the text and a test dataset is used to measure the performance of the classifier [31]. In unsupervised approach, the machine is trained using information that is not labelled or classified. Therefore, the primary task in unsupervised learning is to build a dictionary, used to search and cluster features based on sentiments [8]. A third approach known as semi supervised uses less training data and large volumes of test data. A large number of machine learning problems come under this category, owing to the fact that unlabeled data is easily accessible and cheap as compared to labeled data which is time consuming, expensive and requires expertise.

C. *Deep Learning based Sentiment Analysis*

Deep Learning is a novel method of Machine Learning and it is based on the number of layers in a Neural Network. The main advantage of deep learning is that it extracts features implicitly without giving handcrafted features. Three types of layers are used in Neural Network viz. input, hidden and output layers [9]. Deep learning is a powerful machine learning technique, as deep learning models are more effective in learning features from the text automatically and

produce better results [32]. Some of the deep learning models used in NLP tasks are Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory Network and Recursive Neural Network.

IV. ISSUES IN SENTIMENT ANALYSIS OF MULTILINGUAL TEXT

As we know English is the widely used language on the internet and on the social media. But the hold of general public on the regional languages is strong and these people communicate in their own language especially in a country like India, where a small portion of population knows English. Therefore, the role of regional languages in the social media cannot be undermined. A number of reasons such as deeper connect with audience, increased popularity, business and marketing, and spread of education makes use of local languages favourable on social media. Therefore during the past decade, we see an increase of regional language text on social media platform. However, there are certain issues related to the opinion mining of regional language text data. These include limited volume of labelled dataset, lack of lexical and linguistic tools for languages other than English particularly in Indian languages, complexity and linguistic ambiguity of languages and requirement of content domain methods for better results. [10]

Sentiment analysis or opinion mining from code-mixed data is one of the difficult tasks and the reasons are listed below:

- Generally, code-mixed data is noisy in nature and it requires cleaning and normalization.
- It needs several steps such as language identification and POS tagging.
- There is no sentiment annotated code-mixed lexicon available for any language pairs.
- The available code-mixed datasets are small in size to perform any unsupervised classification

There are a lot of areas that needs to be covered when performing sentiment analysis of code-mixed script; such as, Language Identification, POS Tagging, Sentiment scores generation and training the machine learning algorithm to perform the classification [11].

V. LITERATURE SURVEY

The work in [4] is the first known work for sentiment analysis in Hindi and an early one for an Indian language. The authors use three approaches viz. i.) Construction of a classifier model for Hindi, ii.) Train a model on annotated English corpus and translate Hindi Document to English iii.) Using a majority based classifier for Hindi “*SentiWordNet*”. The results shows that the approach of annotated corpus of Hindi language shows better results (78.14 %) than the other two.

The authors in [12] proposed a method that includes large volume of Hindi “*SentiWordNet*” (HSWN) so that classification yields better results. The approach uses HSWN for word polarity on movie review corpus and achieved an overall accuracy of 80.21%.

A bootstrap approach was proposed in [13] in which a sentiment lexicon for Hindi in the absence of annotated corpus was presented. The proposed approach extracted

sentiment words from Hindi “*WordNet*” to minimize the wrong polarity words. A multi module sentiment analysis system for Hindi using Hindi Sentiment lexicon was used on two domains viz. movie and product. The results showed an average accuracy of more than 85%.

In [14] the authors applied SVM and Decision tree based J48 classifier on a limited trained data of Hindi tweets. To extract sentiments from the tweets weighted unigrams and tweet specific features were exploited. The proposed method achieved an accuracy of 43%.

For forecasting Election Results in India from Hindi language tweets, the authors in [15] proposed a supervised as well as an unsupervised approach using Naïve Bayes, SVM and Dictionary based algorithms for classifications. The results showed that the SVM algorithm provides better results of 78.4% than the other two approaches.

In [16] opinion polarity classification on news text for Bengali language using SVM was presented. The classifier uses sentiment lexicon, POS tags and theme clusters as resources. The SVM-based automatic opinion polarity detection algorithm provides a precision of 70% and a recall of 63%.

The authors in [17] proposed a sentiment analysis for Hindi and Bengali tweets. A multinomial Naïve Bayes classifier is trained on labeled data for classification of tweets into positive, negative and neutral. The proposed method achieved an overall accuracy of 50%.

The authors in [18] used SAIL (Sentiment Analysis of Indian Languages) dataset of sentiment annotated tweets in three Indian languages viz. Hindi, Bengali and Tamil. A 2-class classification and 3-class classification were applied on the data set with four categories of features such as word n-grams, character n-grams surface features and “*SentiWordNet*” features. NB, DT, RF, SVM and LR classifiers were used to achieve an overall accuracy of 51%.

Supervised machine learning algorithms for Tamil movie reviews were presented in [8] SVM, Maxent Classifier, Decision Tree and Naïve Bayes algorithms were used to categories the reviews. Tamil “*SentiWordNet*” is used for higher accuracy. F-score measure was used for classifier examination. The results showed that the SVM performs better.

A Fuzzy Logic Based Hybrid approach for sentiment analysis of Malayalam movie reviews collected from web was presented in [19]. The input is tokenized using “*TnT*” tagger for application of rules to get individual scores of each word. The overall score is provided to a Triangular membership function. The Fuzzy membership function is used to calculate the positive and negative sentiments. The proposed method achieved an accuracy of 91%.

A comparison of lexicon and supervised machine learning based approaches for sentiment analysis of Urdu blogs on multiple subjects is presented in [20]. In supervised learning approach SVM, Decision Tree and K-nearest Neighbor classifiers were applied whereas Urdu Sentiment Lexicon and Urdu Sentiment Analyzer were applied in lexicon approach. Experimental results showed better results in terms of Accuracy, Precision, Recall and F-measure for Lexicon approach.

An overview of the shared task at ICON-2017 on

sentiment analysis of code-mixed data sets of English-Hindi and English-Bengali collected from the different social media platform was presented in [21]. Classifiers such as SVM, Naïve Bayes, were applied on unigram, bigram and n-gram features for sentiment classification.

The authors in [22] presented an analysis Hindi-English code-mixing from Facebook users who express in more than one language. The study reveals that of Hindi in English text and English in Hindi text is noteworthy. It was also revealed that the mixing of Hindi in English follows a particular pattern but the mixing of English in Hindi can occur at any place. The classification of Code-Mixed words based on frequency and linguistic typology shows that a large majority of the words form a group in the middle, therefore needs to be handled at multiple levels for automatic processing of the data.

The POS (Part of Speech) tagging of social media text of three Indian languages (Hindi, Bengali and Tamil) was addressed in [23]. Two approaches for POS tagging such as POS tagging using Stanford POS tagger and POS tagging using CRF++ were used for training data and model construction. Results showed that the language pair Tamil-English has less accuracy than English-Hindi and English-Bengali language pairs because of lack of quality training data. The English-Bengali Code-mixed data has highest accuracy of 75%.

An English-Hindi code mixed corpus from social media platform (Facebook) was created and analyzed in [24]. The corpus created from public pages of three Indian celebrities has 6983 posts and 113578 words. The analysis include language identification, transformation and POS tagging. The results show that normalization and transformation for English-Hindi code mixed text are two major challenges that impact POS tagging accuracy.

The method proposed in [25] used SVM with radial basis function for 2866 Hindi-English code-mixed tweets annotated with six emotions and achieved an accuracy of 58.2%.

Sentiment analysis of the “*Hinglish*” text by using dictionary based technique was proposed in [26]. “*Hinglish*” movie review dataset was collected for sentiment analysis. The authors prepare two dictionaries: one for English data and another for Hindi data which are capable of handling word variations and are case-insensitive. For feature extraction, *tf-idf* technique was used along with unigram, bigram and trigram. For sentiment classification, SVM, Naïve Bayes, Neural Network and Logistic regression algorithms were used for identification of best feature set and classifier for “*Hinglish*” text.

Authors in [27] presented a language tagged and POS-tagged dataset of code-mixed English-Hindi tweets related to five incidents in India that led to a lot of Twitter activity. The dataset created was larger than previous annotated datasets and closely resembles typical real-world tweets. A POS tagging model that is trained on this dataset to provide an example of how this dataset can be used was also presented. The model shows the efficacy of dataset in enabling the creation of code-mixed POS taggers for social media text.

Authors in [28] proposed a feature based user review polarization and classification of Hindi text. The annotated

dataset containing 5417 user reviews was created from online reviews of different computer, mobile, travel and movie domains. To train multi-labeled classifier n-grams, character n-grams, bigram and trigram features were used. Supervised learning algorithms such as Naïve Bayes, SVM were used and the results showed that Naïve Bayes performs better. For aspect category detection, an average F-measure of 50% and for sentiment classification, the proposed model achieved an overall accuracy of about 67%.

Sentiment analysis of code mixed bilingual text (Hindi-English) obtained from social media sites like Facebook was proposed in [29]. Two different approached for code mixed Indian language text were used. A lexicon of movie posts was constructed which contains code mixed sentences. From these sentences sentiments were extracted and a model was trained. A polarity detection algorithm was applied to obtain the overall sentiment of a sentence. A second method based on machine learning approach unigrams, bigrams and POS tag features were used to create a feature vector and train a model. For sentiment classification SVM, Naïve Bayes, Decision Tree and Random Tree were applied. Experimental results showed that the lexicon approach yields better accuracy.

A comparison between various Machine Learning methods to perform polarity classification of tweeter data in multilingual environment viz. single language text, text in different languages and code-mixed text was proposed in [30]. Three machine learning models viz monolingual, multilingual model and monolingual pipe model and extracted features were evaluated on monolingual, multilingual and code-mixed corpus. The experimental results on these three different types of corpus showed that accuracy obtained on monolingual corpora was higher than the code-switching corpus.

VI. CONCLUSION

Sentiment Analysis on code mixed languages is an evolving research area that has received a lot of attention in recent years. As already discussed, Sentiment Analysis of text involves a number of stages starting from collection of data, preprocessing, detection and classification of sentiment, to the presentation of results. In this paper, an attempt has been made to study the work done in the field of Natural Language processing and sentiment analysis of Indian language texts on social media and microblogging sites in the form of opinions, emotions, feelings, reviews and suggestions. The study shows that most of the work has been done in Hindi, Bengali and Tamil languages including 70% of research that has been carried out on Indo-Aryan Languages. As the regional Indian language data on the social media sites increases, the work on other languages has also started. However, due to the unavailability of lexicon tools and corpus in most of the languages, the researchers find it quite challenging to carry out their work. In the recent past the research focus also shifted to code-mixed and code-switched Indian languages as people knew more than one language and express their sentiments in more than one language. Although a significant work has been carried out in Hindi and other Indian language text, the work on code mixed Indian language data is also showing positive trends. As a part of future work, we will explore and evaluate existing code-mixed data sets and corpus using

traditional machine learning approaches. Also, we will work towards proposing a novel approach for detecting sentiments in code mixed Indian languages using machine learning techniques.

REFERENCES

- [1] Rani, Sujata, and Parteek Kumar. "A sentiment analysis system to improve teaching and learning." *Computer* 50, no. 5 (2017): 36-43.
- [2] Rani, Sujata, and Parteek Kumar. "A journey of Indian languages over sentiment analysis: a systematic review." *Artificial Intelligence Review* (2018): 1-48.
- [3] Kaur, Jasleen, and Jatinderkumar R. Saini. "A study and analysis of opinion mining research in Indo-Aryan, Dravidian and Tibeto-Burman Language families." *International Journal of Data Mining and Emerging Technologies* 4, no. 2 (2014): 53-60.
- [4] Joshi, Aditya, A. R. Balamurali, and Pushpak Bhattacharyya. "A fall-back strategy for sentiment analysis in hindi: a case study." *Proceedings of the 8th ICON* (2010).
- [5] Asghar, Muhammad Zubair, Aurangzeb Khan, Syeda Rabail Zahra, Shakeel Ahmad, and Fazal Masud Kundi. "Aspect-based opinion mining framework using heuristic patterns." *Cluster Computing* (2017): 1-19.
- [6] Rehman, Zia Ul, and Imran Sarwar Bajwa. "Lexicon-based sentiment analysis for Urdu language." In *Innovative Computing Technology (INTECH)*, 2016 Sixth International Conference on, pp. 497-501. IEEE, 2016.
- [7] Syed, Afraz Z., Muhammad Aslam, and Ana Maria Martinez-Enriquez. "Lexicon based sentiment analysis of Urdu text using SentiUnits." In *Mexican International Conference on Artificial Intelligence*, pp. 32-43. Springer, Berlin, Heidelberg, 2010.
- [8] Se, Shriya, R. Vinayakumar, M. Anand Kumar, and K. P. Soman. "Predicting the sentimental reviews in Tamil movie using machine learning algorithms." *Indian Journal of Science and Technology* 9, no. 45 (2016).
- [9] Seshadri, Shriya, Anand Kumar Madasamy, Soman Kotti Padannayil, and M. Anand Kumar. "Analyzing sentiment in indian languages micro text using recurrent neural network." *IIOAB J 7* (2016): 313-318.
- [10] Wang, Zhaoxia, Victor Joo, Chuan Tong, and David Chan. "Issues of social data analytics with a new method for sentiment analysis of social media data." In *Cloud Computing Technology and Science (CloudCom)*, 2014 IEEE 6th International Conference on, pp. 899-904. IEEE, 2014.
- [11] Ansari, Mohammed Arshad, and Sharvari Govilkar. "SENTIMENT ANALYSIS OF MIXED CODE FOR THE TRANSLITERATED HINDI AND MARATHI TEXTS." *International Journal on Natural Language Computing (IJNLC)* (2018)
- [12] Mittal, Namita, Basant Agarwal, Garvit Chouhan, Nitin Bania, and Prateek Pareek. "Sentiment Analysis of Hindi Reviews based on Negation and Discourse Relation." In *Proceedings of the 11th Workshop on Asian Language Resources*, pp. 45-50. 2013.
- [13] Sharma, Raksha, and Pushpak Bhattacharyya. "A Sentiment Analyzer for Hindi Using Hindi Senti Lexicon." In *Proceedings of the 11th International Conference on Natural Language Processing*, pp. 150-155. 2014.
- [14] Venugopalan, Manju, and Deepa Gupta. "Sentiment Classification for Hindi Tweets in a Constrained Environment Augmented Using Tweet Specific Features." In *International Conference on Mining Intelligence and Knowledge Exploration*, pp. 664-670. Springer, Cham, 2015.
- [15] Sharma, Parul, and Teng-Sheng Moh. "Prediction of indian election using sentiment analysis on hindi twitter." In *Big Data (Big Data)*, 2016 IEEE International Conference on, pp. 1966-1971. IEEE, 2016.
- [16] Das, Amitava, and Sivaji Bandyopadhyay. "Phrase-level polarity identification for Bangla." *Int. J. Comput. Linguist. Appl.(IJCLA)* 1, no. 1-2 (2010): 169-182.
- [17] Sarkar, Kamal, and Saikat Chakraborty. "A sentiment analysis system for Indian language tweets." In *International Conference on Mining Intelligence and Knowledge Exploration*, pp. 694-702. Springer, Cham, 2015.
- [18] Phani, Shanta, Shibamouli Lahiri, and Arindam Biswas. "Sentiment analysis of tweets in three Indian languages." In *Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP2016)*, pp. 93-102. 2016.
- [19] Anagha, M., Raveena R. Kumar, K. Sreetha, and PC Reghu Raj. "Fuzzy logic based hybrid approach for sentiment analysis of Malayalam movie reviews." In *Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, 2015 IEEE International Conference on, pp. 1-4. IEEE, 2015.
- [20] Mukhtar, Neelam, Mohammad Abid Khan, and Nadia Chiragh. "Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains." *Telematics and Informatics* 35, no. 8 (2018): 2173-2183.
- [21] Patra, Braja Gopal, Dipankar Das, and Amitava Das. "Sentiment Analysis of Code-Mixed Indian Languages: An Overview of SAIL_Code-Mixed Shared Task@ ICON-2017." *arXiv preprint arXiv:1803.06745* (2018).
- [22] Bali, Kalika, Jatin Sharma, Monojit Choudhury, and Yogarshi Vyas. "'I am borrowing ya mixing?'" *An Analysis of English-Hindi Code Mixing in Facebook.* In *Proceedings of the First Workshop on Computational Approaches to Code Switching*, pp. 116-126. 2014.
- [23] Ghosh, Souvick, Satanu Ghosh, and Dipankar Das. "Part-of-speech tagging of code-mixed social media text." In *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, pp. 90-97. 2016.
- [24] Vyas, Yogarshi, Spandana Gella, Jatin Sharma, Kalika Bali, and Monojit Choudhury. "Pos tagging of english-hindi code-mixed social media content." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 974-979. 2014.
- [25] Vijay, Deepanshu, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. "Corpus Creation and Emotion Prediction for Hindi-English Code-Mixed Social Media Text." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pp. 128-135. 2018.
- [26] Kaur, Harpreet, Veenu Mangat, and Nidhi Krail. "Dictionary based Sentiment Analysis of Hinglish text." *International Journal of Advanced Research in Computer Science* 8, no. 5 (2017).
- [27] Singh, Kushagra, Indira Sen, and Ponnurangam Kumaraguru. "A Twitter Corpus for Hindi-English Code Mixed POS Tagging." In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pp. 12-17. 2018.
- [28] Akhtar, Md Shad, Asif Ekbal, and Pushpak Bhattacharyya. "Aspect Based Sentiment Analysis: Category Detection and Sentiment Classification for Hindi." In *International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 246-257. Springer, Cham, 2016.
- [29] Pravalika, A., Vishvesh Oza, N. P. Meghana, and S. Sowmya Kamath. "Domain-specific sentiment analysis approaches for code-mixed social network data." In *Computing, Communication and Networking Technologies (ICCNT)*, 2017 8th International Conference on, pp. 1-6. IEEE, 2017.
- [30] Vilares, David, Miguel A. Alonso, and Carlos Gómez-Rodríguez. "Sentiment analysis on monolingual, multilingual and code-switching twitter corpora." In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 2-8. 2015.
- [31] Jain, Anuja P., and Padma Dandannavar. "Application of machine learning techniques to sentiment analysis." In *Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2016 2nd International Conference on, pp. 628-632. IEEE, 2016.
- [32] Zhang, Lei, Shuai Wang, and Bing Liu. "Deep learning for sentiment analysis: A survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (2018): e1253.
- [33] Montoyo, Andrés, Patricio MartíNez-Barco, and Alexandra Balahur. "Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments." (2012): 675-679.