International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 1656–1675

Recent Advances in Sentiment Analysis of Indian Languages

Mahesh B. Shelke[1], Sachin N. Deshmukh[2]

Department of Computer Science & Information Technology,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (M.S.), India
[1]mahesh_shelke21@hotmail.com
[2]sndeshmukh@hotmail.com

Abstract

In the era of technology, each and every one of us is expressing their opinion on social media platforms very frequently. And these opinions are mostly expressed in regional languages, so the contents mostly generated are in regional languages in nature. Sentiment Analysis (SA) is a natural language processing task that is defined as finding opinion (In the sense of Positive, Negative, or Neutral) of the writer about specific entities. This includes analyzing a person's emotions, feelings, and attitudes towards his contents. This paper gives a comparative analysis of sentiment analysis performed in various Indian languages, which includes classification techniques which are based on Lexicon, Dictionary, and Machine Learning. And it also gives a list of lexical resources available to perform Sentiment Analysis (SA) of Indian Languages and the challenges of developing lexical resources for low resourced Indian languages.

Keywords: *Sentiment Analysis, Indian Languages, Lexical Resources, Classification techniques, Machine Learning.*

1. Introduction:

In world of digitalization millions of people are connected to each other through World Wide Web and social networking. This allows new way of sharing contents to other users. Social Networks, E-commerce website, blogs etc. are different ways, which allows users to generate and share their contents, ideas and opinion with other easily, which leads to generate a huge amount of data every day.

Sentiment analysis and opinion mining emerged as a challenging and active field of research for resourced as well as low resourced languages. The term opinion covers the broad concept that is Sentiment, evaluation, appraisal or attitude of piece of information which indicates the opinion of writer or speaker.

These opinion of users varies from user to user. So it is necessary to consider several number of opinion uses which will give more accurate sentiments related to the topic, so it becomes necessary to analyze large amount of opinion.

Sentiment are associated with three terms: type, orientation and intensity. These types are based on linguistic, psychological and consumer research. Consumer research based sentiment having two type as rational and emotional sentiment. Rational sentiments depends on rational reasoning, tangible beliefs and utilitarian attitudes. As "*The camera of the phone is good" and "This phone is worth the price"* indicates Rational Sentiments. And Emotional sentiments are based on psychological state. As "I love my iPhone" indicates the emotions towards the phone.

Sentiment orientation or polarity is indicated as positive, negative and neutral. And Intensity is indicated as how user expresses his opinion, for example *#1:*"Camera of this phone is excellent" and *#2:* "Camera of this phone is good". If we compare both the sentences, we may find that opinion which includes the word *"excellent"* is stronger than opinion which includes the word *"good"*. Same as *"detest"* is stronger that

ISSN: 2233-7857 IJFGCN

"*dislike"*. Intensity of sentiment also depends on the some words like *very, so, extremely, really, awful, terrible, slightly, pretty* etc. which allows author to express his opinion in strong manner [18].

1.1 Indian Language Families

India is a multilingual country with 22 official languages spoken in India. These languages belongs to several families as Indo-Aryan (Arya), Dravidian (Dravida), Austroasiatic (Nishada), Sino-Tibetan (Kirta) and few other minor languages. Out of this, Indo-Aryan languages are spoken by 78.05% and the Dravidian languages spoken by 19.04% of Indian Population [17].

Indo-Aryan family consist of languages as Hindi, Urdu, Bengali, Oriya, Punjabi, Konkani, Marathi, Nepali, Gujarati, Sindhi, Dogri, Assemese, Sanskrit and Kashmiri. Dravidian family consist of languages such as Telugu, Tamil, Malayalam, and Kannada. Austroasiatic family consist of Santali as official language and Sino-Tibetan language family includes languages as Manipuri and Bodo [42].

1.2 Need for Sentiment Analysis of Indian Languages

- As Indian Government has launched Digital India initiative which allows rapid growth of Internet Users, as of January 2020 around 688 million active users, which makes India as second highest online internet market after China [11].
- A new era of internet users in the India are opting to access the internet in their Regional language. And by 2021, there will be around 500 million users of Indian Languages while English users will be less than half [9]. Soon English speaking user base will be overtaken by Hindi speakers from India, and Hindi will become most used language over internet in India. And other languages like Marathi, Bengali, Tamil, and Telugu will share 30% of internet users in India.
- Most of the Search Engines, Android/IOS Application, Social Networking Platforms, and Government Websites are now a days available in Indian Languages, and this generates a huge amount of data over Internet. Which allows researcher to explore the research field in Natural language processing for Indian languages. And Sentiment Analysis is one of the major task of NLP, so it is important to develop the resources for Indian languages to perform Sentiment Analysis (SA) on Indian Languages (ILs).

1.3 Motivation of Research

From the above, it can be conferred that some Indian languages like Hindi, Bengali, Tamil, Malayalam and Telugu have been considered for performing sentiment analysis and accordingly the linguistic and lexical resources has been developed for these languages, but still there are some languages like Marathi, Urdu, Punjabi etc. which are not been considered for Natural language Processing task of SA to greater extent. For this task, it becomes important to perform systematic review for Indian languages for which it is already developed, and it may be beneficial for other Indian Languages (ILs) for further implementation.

1.4 Our Contributions

We have performed systematic review of Indian languages (ILs) for which implementation of Sentiment Analysis (SA) task is already completed. These languages are Hindi, Bengali, Tamil, Malayalam, Telugu and Konkani. This review article considers research published in Journal, Conferences, and Workshops since 2015. And also revealed available linguistic and lexical resources, annotated data sets which may be helpful for future work related to Sentiment Analysis (SA). A survey of various sentiment classification techniques, related work done, past literature, list of available linguistic and lexical resources and challenges associated to perform Sentiment Analysis (SA) of Indian Languages (ILs) is also presented here.

2. Literature Survey

ISSN: 2233-7857 IJFGCN

Sentiment Analysis became one of the prominent field for Researchers since last decade. But still very few languages like English, Chinese, Hindi, Arabic etc. has been majorly explored. And some languages are still unexplored in the field due to various challenges of languages like lack of lexical and linguistic resource.

As far as Indian Languages are concerned, some languages like Hindi, Bengali, Tamil, Telugu, Malayalam has been majorly explored but still some of them are unexplored or minor work has been done on languages such as Marathi, Punjabi, Konkani, Dogri, Gujarati, Kannada, Oriya, etc.

Authors have developed framework for Sentiment Analysis (SA) using HindiSentiWordNet (HSWN), which uses Synset Replacement Algorithm for replacing Synset words which is having closest meaning word available in HSWN, for finding polarity of words which are not available in HSWN [29]. Authors have proposed system for Sentiment Analysis of Hindi Tweets using unsupervised classification, which uses subjective lexicon. Authors have developed SentiWordNet which includes adjectives and adverbs with highest accuracy as 81.97% [44].

Hindi Opinion Mining System (HOMS) proposed by authors, which classifies documents using supervised (Naïve Bayes) and unsupervised(POS Tagging) methods. After POS tagging, adjectives are considered as words which expresses opinion and Adjective word count is used to classify documents. After Negation Handling, unsupervised method gives highest accuracy as 91.4% and NB gives highest accuracy as 87.1% [15]. Authors have developed resources for performing aspect based SA of Hindi. In which benchmark annotated dataset has been developed, and every sentence of the review is marked with aspect term and its sentiment and for classification they have used SVM and CRF for aspect term extraction. This system has limitation as if sentiment words are present far away from the aspect term then system does not generate correct polarities [24].

Authors have developed system for Multiclass classification and class based sentiment analysis (SA) for Hindi language, in which authors have used Hindi-SentiWordNet (HSWN) with language model (LM) classifier to increase the word coverage. This system has limitation that test document should contain min. 500 and max. 100 words in it. And they have implemented static ontology instead of self-learning ontology [31]. Authors have developed system for prediction of Indian election using Hindi tweets, using supervised (NB, SVM) and unsupervised (Dictionary based) approaches for classification. For better results Hindi SentiWordNet need to be improved with more number of word. It contains synonym and antonym words with sentiment score [30].

Authors have proposed aspect based sentiment classification for Hindi. They have created annotated dataset for aspect category detection and classification of sentiment. This system is tested on domains as Electronics, mobiles apps, travels and movies [3]. Authors has proposed method which uses domain knowledge to perform sentiment analysis using ontology. As Hindi documents are having lack of words in it, they have created static ontologies. And the accuracy of the system depends on size of the dataset used for training and testing the Model [8].

Authors have proposed different combinations of n-gram and SentiWordNet features for sentiment classification of Bengali language. MNB and SVM is used for classification in which SVM gives better accuracy than MNB with these features. And the performance of system depends on size of training dataset and how well human annotators have labeled dataset [16]. Authors have proposed method finds the polarity of Bengali tweets using LSTM. However Performance of system depends on correctly labeled dataset, if datasets are wrongly labeled which results in problem of over-fitting as on model has effect of dropout [36]. Authors have proposed system which is based on supervised classification which is improved using lexicon expansion approach as Distributional Thesauri and Co-occurrences which gives the highest accuracy as 49.68% and 43.20% for Hindi and Bengali respectively with constrained [7].

Supervised machine learning algorithm as MNB, BNB, RKS and SVM. In which they have considered frequency of occurrences of keywords as suitable feature for performing Tamil movie sentiment analysis.

ISSN: 2233-7857 IJFGCN

Results indicates that SVM has outperformed with highest accuracy as 64.6% [6]. Authors have developed system for prediction of sentimental reviews in Tamil movie using machine learning algorithm as SVM, Maxent, DT and NB. In which SVM outperforms other classifier with highest accuracy [38]. Proposed model for sentiment classification of Tamil movie tweets uses TF-IDF, TF-IDF+DST(Domain Specific Tags) and Tweet weightage model. In which tweet weightage model gives better accuracy than others. But the Performance of Tweet weightage model depends on sentiment lexicon of that particular domain [26]. Authors have proposed system which focuses on long, ambiguous and sarcastic phrases. This system performs sentiment analysis of online Tamil contents using RNN. This system does not implements sarcastic phrase detection [28].

Authors have developed sentiment mining system for Tamil and Bengali tweets based on probabilistic(NB) and Decision tree(C4.5) for classification. And the Performance of system depends on size of dataset used, many variations in words, and improper use of punctuation marks [35].

Proposed system of sentiment analysis(SA) which considers three Indian languages Bengali, Hindi and Tamil. It considers features as n-grams, surface and SentiWordNet. This system do not have access to Tamil test data and stop words [37]. Sentiment analysis(SA) of three Indian languages as Hindi, Tamil and Bengali using Recurrent Neural Network (RNN). As RNN is having problem of long term dependency which can be overcome using LSTM [39].

Authors have developed Fuzzy logic based Hybrid approach for performing SA of Malayalam movie reviews. Which includes Machine Learning algorithm for Tagging and to find membership of reviews using fuzzy logic. And TnT Tagger is used to train manually tagged dataset [4]. Authors have performed sentiment analysis of Malayalam twitter dataset using Rule-Based approach. Polarity of the document is calculated based on positivity and negativity values of individual documents [23].

Proposed system is multimodal (Text and Audio) sentiment analysis(SA) of Telugu songs which uses lyrics feature which are created using doc2vec and for audio they have considered features as spectral, Chroma etc. [12]. Proposed system uses the Rule-Based and Machine Learning based approaches for sentiment analysis of Telugu movie reviews and also Authors have created seed list of words [40].

Authors have proposed system which uses Gate Processor(NLP) for sentiment classification of Marathi language. As the performance of the system mostly depends on text [10]. Proposed method focuses on translitered script of Hindi and` Marathi language. They have used bilingual(Hindi and Marathi) dictionary as Hindi-SentiWordNet(HSWN) and English- SentiWordNet [25]. Authors have developed system for sentiment analysis of Marathi language using WordNet. And this system mainly focuses on creation of dataset, mapping of dataset and performing accurate classification. This system translates Marathi text into English using Yandex translator and maps it into English SentiWordNet. And System performance depends on how many number of words are actually covered in WordNet and size of dataset used [41].

Model for sentiment analysis of Konkani text. Which includes 50 poem with at least one sentiment tag present in corpus and 10 randomly selected poems which results in 82% and 70% accuracy. This model need to be fined tuned to perform as like human annotators [5]. Authors has proposed approach for sentiment analysis(SA) of Gujarati tweets, in which they have used POS tagging to extract features and SVM for classification. They have used only 40 tweets as sample dataset [43].

Proposed model uses Baseline, Lexicon and Naive Bayes Model for analyzing code mixed text. In this syntactic rules plays important role while implementing code switching [22]. Authors have proposed system for sentiment extraction from bilingual (English-Telugu) text, which performs language identification, back transliteration to native language and classifying the sentiment of text [27]. Proposed method is the strength of rich sequential pattern of LSTM and words from n-gram probabilistic model(MNB) for code mixed data i.e. Hindi-English to identify sentiment [21].

ISSN: 2233-7857 IJFGCN

Proposed method uses Recursive Auto Encoder Architecture to develop CLSA Tool between pair of resourced(English) and low-resourced(Hindi) languages. This system works on semantic similarity between phrases and sentiment similarity between sentences. With Basic RAE, BRAE-U, BRAE-P, BRAE-F Machine Translation system can be reduced using lexical resources without affecting the performance [14].

3. Sentiment Analysis

Finding the contextual polarity of people's sentiment's, opinion, attitudes, appraisal and emotions in text is the task of Sentiment Analysis (SA). It gives polarity of text in terms of Positive, Negative and Neutral. Mainly researchers have studied sentiment analysis at different levels as Document level, Sentence level and Aspect level sentiment classification. In Document level sentiment classification is done on whole documents which results in document may be Positive, Negative or Neutral. In Sentence level classification individual sentences indicates polarity as Positive, Negative or Neutral. And Aspect level classification considers aspect terms present in sentences which results in opinion related to aspect terms, Aspect level sentiment analysis is divided into different task as extraction of aspect, extraction of entity and sentiment classification of aspect.

3.1 Techniques of Sentiment Classification

Sentiment classification techniques are categorized as Machine Learning Approach, Lexicon Based and Hybrid Approach. In Machine learning approach are divide into supervised and unsupervised learning. Supervised learning uses large labeled dataset for training the model and whereas unsupervised learning is used when there is no labeled dataset are available to train the model.

Lexicon based approaches are mostly depends on lexical resources which indicates the opinion of text. It is based on Corpus based method and dictionary based, In corpus based method domain specific seed list of opinion words are included which helps to find the polarity of text and In dictionary based method a small set of opinion words are collected and it is manually annotated, which includes synonyms and antonyms and also it can be expanded using WordNet, Distributional Thesaurus and sentence level co-occurrences [7]**,** and Senti2Vec [20]**.** Hybrid approach combines the Machine learning approach and lexicon based approach with sentiment lexicon.

3.1.1 Machine Learning based approach

Mostly in machine learning based method uses several Machine Learning algorithms to perform SA as a text classification problem which makes the use of linguistic and syntactic features.

A. Supervised learning

In supervised learning two terms are frequently used as training dataset and testing dataset. Training dataset are mostly used by classifier to train the model with different characteristics associated with documents and find the accuracy of the classifier with test dataset. It goes through the two stage process as Learning the Model using training dataset and testing the model using test dataset. There are various supervised learning classifier as Probabilistic, Rule-Based, Decision Tree and Linear Classifiers.

a. **Probabilistic classifiers** are Naïve Bayes Classifier(NB), Bayesian Network(BN), and Maximum Entropy Classifier(ME). Naïve Bayes Classifier is based on Bayes theorem, which is used to find posterior probability of class, based on the words are distributed in the particular text documents. And documents are classified using Bag of Words feature extraction, while in document position of word is not considered in classifications. And Model uses Bayes theorem to predict probability that a given feature belong to particular class. As Bayesian Network's(BN) computation complexity is

ISSN: 2233-7857 IJFGCN

very expensive that is why it not used frequently. And Maximum Entropy classifier is also known as a conditional exponential classifier, which creates encoded vectors by converting labeled features sets. These vectors are used to calculate weights of each feature that can be merged to define most likely label for feature set.

b. **Rule-Based Classifier** uses different set of rules for generating opinion, which can be created by performing tokenization of each sentence present in documents and finally it is tested for its presence. Rules are created using various criteria which are used to generate rules. Example, if available word or token indicates positive sentiment then it applies +1 ratings. If final polarity score of sentence is greater than zero which indicates positive sentiment else if score is less than zero it indicates negative sentiment.
c. **Decision Tree(DT)** classifier uses hierarchical decomposition of training dataset in which condition on the attribute value is used to split the data. Presence or absence of one or more words indicates predicate or condition. Dataset is divided recursively till leaf nodes which contains minimum number of records for classification.
d. **Linear Classifier** defines the class of an object to which it belongs by using classification decision based on the value of a linear combination of objects with its characteristics. And classification is based on combination of their weights and features. Support Vector Machine (SVM) is a linear classifier which works on concept of hyperplane and decision plane. SVM performs classification by defining the best hyperplane, which differentiate two different classes. By training the model to classify dataset using training dataset, which allows to define best hyperplane. Once we have trained model and best hyperplane we can classify any dataset.

B. Unsupervised

Mostly unsupervised learning is preferred when unlabeled dataset are available. It works with defining hidden structure present in unlabeled datasets. K-Nearest Neighbor (KNN) ML algorithm is used in unsupervised learning, which classifies the datasets using nearest neighbor of the objects and assign it a particular class.

3.1.2 Lexicon Based approach

In sentiment analysis the word which express the opinion are most important because it signifies the sentiment about text in terms of positive, negative and neutral or indicates the polarity score. In lexicon collection of words with its associated polarity score are predefined. Lexicon are used to perform word matching in terms of categorize a sentence. And performance of lexicon based approach depends on size of lexicon. For example, if any document contains more positive words from lexicon than negative words, then that document is positive. There are two methods under lexicon based approach as dictionary based and corpus based method. These methods are used to create sentiment lexicon. And these lexicon are manually constructed with the help of humans to manually assign polarities to words which is difficult and time consuming task.

As Dictionary based method is an iterative process, in which initially small set of seed list is manually created, which contains sentiment words. Then these seed list grows by adding synonyms and antonyms of seed words. This seed list grows until there is no word left for adding in sentiment lexicon. There are various dictionaries are available like WordNet, SenticNet, SentiWordNet, Hindi-SentiWordNet etc.

In corpus based methods, manually seed list of words is constructed and then by using syntactic pattern of these seed list words is used to construct new words from the large corpora. Syntactic patterns refers to co-occurrences of words or how frequently words appears with each other. Mostly this method is used to generate domain specific opinion words.

3.2 Comparison of different sentiment analysis Techniques used in Indian Languages

ISSN: 2233-7857 IJFGCN

Table 1 gives the comparative analysis of Indian Languages which are used to perform Sentiment analysis. Which includes language, classification techniques, lexicon type, dataset size, dataset, tools used and evaluation measures.

4. Resources available for Indian Languages

There are some languages having linguistic and lexical resources which have been developed such as Hindi, Bengali, Tamil, Telugu, and Malayalam. These resources are developed by using supervised and unsupervised techniques such as Distributional Thesaurus and Co-occurrences (DT_COOC), Building lexicon manually using human annotators, using Machine Translation Techniques etc. There are some resources required for preprocessing such as Shallow Parser, Morph Analyzer, POS Tagger, CRF Chunker [1]. IndoWordNet [13]. And SentiWordNet [2][5]. Available for some languages. Following *Table 2* shows available NLP resource for Indian Languages.

Table 1: Comparative analysis of diff. techniques used in Sentiment Analysis of Indian Languages;

Reference	[7]	[6]	[29]	[14]	[4]
Language	Hindi, Bengali	**Tamil**	Hindi	**Hindi-English**	Malayalam
Technique Used	SVM	**Multinomial and Bernoulli Naive Bayes, Logistic Regression, SVM, Random Kitchen Sink**	Synset Replacement Algorithm	**BRAE Framework**	Fuzzy logic
Lexicon Type	Microblogging	**Reviews Collected from 25+ websites**	Microblogging, Reviews	**HindMonoCorp 0.5, and Movie Reviews**	Reviews Collected from websites
Dataset Size	Not Specified	**2320 Movie Reviews (1160 - +ve and –ve)**	Not Specified	**HindMonoCorp 0.5 with 44.49M sentences and English Gigaword Corpus and IMDB11 Contain 25000 +ve and 25000 –ve movie reviews**	Training Dataset of 2500 words
Dataset	Hindi Tweets	**Tamil Movie Reviews**	Tweets, Movie Reviews and Blogs	**HindMonoCorp 0.5, IMDB11 Movie Review dataset**	Malayalam Movie Reviews
Resource used	Hindi and Bengali SentiWordNet	**Machine Learning Algorithms**	Hindi-SentiWordNet	**WordNet, Bag-of-words**	TnT Tagger

ISSN: 2233-7857 IJFGCN

Evaluation Measures	**Hindi** **A:** 49.68% (Constrained) **A:** 46.25% (Unconstrained) **Bengali** **A:** 43.20% (Constrained) **A:** 42% (Unconstrained)	**SVM -** **A:** 64.69% (Bigram) **MNB-** **A:** 47.21% (Bigram)	Not Measured	**RHMR-Ratings** BASIC RAE **A:** 75.53% BRAE-U **A:** 76.01% BRAE-P **A:** 79.7% BRAE-F **A:** 81.22% **RHMR-Polarity** BASIC RAE **A:** 79.31% BRAE-U **A:** 82.66% BRAE-P **A:** 84.85% BRAE-F **A:** 90.5% **SMRD-Polarity** BASIC RAE **A:** 81.06% BRAE-U **A:** 84.83% BRAE-P **A:** 87% BRAE-F **A:** 90.21%	**P:** 91.06%
Observation	Classification performance can be improved using lexicon expansion approach as Distributional Thesauri and Co-occurrences	**Can consider frequency of word as feature.**	Synset Replacement Algorithm can be used for improving sentiment lexicon.	**Considered Semantic and sentiment similarity between phrases and sentences respectively. And also used lexical resources over Machine Translation approach.**	Can be analyzed using more domain specific results for performance improvement

Table 1: Continued…

Reference	[44]	[15]	[24]	[30]	[31]	[25]
Language	**Hindi**	Hindi	**Hindi**	Hindi	**Hindi**	Hindi, Marathi
Technique Used	**Lexicon based**	Naive Bayes Classifier	**CRF for Aspect Extraction and SVM for Classification**	Dictionary Based, Naive Bayes and SVM algorithm	**Lexicon Based, LMC Classifier**	Lexicon Based, SVM, Random Forests

ISSN: 2233-7857 IJFGCN

Lexicon Type	**Microblogging**	Reviews Collected from websites	**Product Reviews, Newspaper, blogs Collected from websites**	Microblogging	**Author Dataset**	Not Specified
Dataset Size	**50 Hindi Tweets of each #जयहिंद and #worldcup2015**	200 Reviews with 80k words	**5417 reviews from 12 domains, 2290 +ve, 712 –ve, 2226 Neutral, 189 Conflict reviews**	42,235 tweets with #tag Political Party name of election 2016.	**Not Specified**	Not Specified
Dataset	**Hindi Tweets**	Movie Reviews	**Product Reviews**	Hindi Tweets related to Political party in India during election 2016	**Hindi speeches delivered by leaders**	Not Specified
Resource used	**Hindi-SentiWordNet**	TnT Tagger	**Shallow Parser**	Hindi-SentiWordNet	**Hindi-SentiWordNet**	SentiWordNet, Hindi-SentiWordNet, POS Tagger
Evaluation Measures	**#जयहिंद – Proposed** **A:** 73.53% **P:** 0.93, **R:** 0.89 **#worldcup2015– Proposed** **A:** 81.97% **P:** 0.77 **R:** 0.87	**After Negation Handling** +ve docs **A:** 91.4% -ve docs **A:** 82.8% **Overall Accuracy:** 87.1%	**SVM:** **A:** 54.05%, **P:** 91.96%, **R:** 30.72% **F:** 41.07%	**Naive Bayes** **A:** 62.1% **P:** 0.71 **R:** 0.61 **SVM:** **A:** 78.4% **P:** 0.75 **R:** 0.78 **Dictionary Based** **A:** 34%	**Not Measured**	Not Measured
Observation	**Lexicon can be improved with more number of word coverage.**	Machine Learning and POS tagging is used to classify documents	**Implemented Aspect Term Extraction and Sentiment classification. And it is hard to detect multi-word aspect terms.**	Emotions associated with aspect are not considered and Dataset is manually labelled.	**Test document size is between 500 to 1000 words only.**	Language coverage can improve better performance of sentiment classification.

ISSN: 2233-7857 IJFGCN

Table 1: Continued…

Reference	[35]	[38]	[37]	[39]	[12]
Language	Bengali, Tamil	**Tamil**	Bengali, Hindi, Tamil	**Tamil, Hindi, Bengali**	Telugu
Technique Used	Naive Bayes and C4.5 Decision Tree	**SVM, Maxent classifier, Decision tree and Naive Bayes**	Lexicon Based, MNB, LR, DT, RF, SVM SVC (SV), and Linear SVC(LS)	**RNN**	Gaussian Mixture Models (GMM), SVM, NB
Lexicon Type	Microblogging	**Reviews Collected from websites**	Microblogging	**Microblogging**	Movie Songs and Lyrics
Dataset Size	999-Bengali tweets and 1103-Tamil tweets	**534 - Tamil Reviews from webpages**	SAIL-2015 Tweets Dataset	**Tamil- 1663, Hindi-1673, Bengali-1499 Tweets**	300 Telugu movie songs and lyrics
Dataset	Bengali and Tamil Tweets	**Tamil Movie Reviews**	SAIL 2015 Tweets Dataset- Tamil, Hindi, and Bengali	**SAIL 2015 Tweets Dataset - Tamil, Hindi, and Bengali**	Telugu movie songs and lyrics
Resource used	WEKA	**Tamil - SentiWordNet**	Bengali, Hindi Tamil SentiWordNet	**Not Used**	doc2vec
Evaluation Measures	**Bengali - Naïve Bayes** Neg -**P:** 0.77 **R:** 0.81 **F:** 0.79 Neu -**P:** 0.70 **R:** 0.81 **F:** 0.75 Pos –**P:** 0.75 **R:** 0.39 **F:** 0.52 **Bengali - C4.5 Decision Tree** Neg - **P:** 0.87 **R:** 0.88 **F:** 0.88 Neu - **P:** 0.75 **R:** 0.87 **F:** 0.81 Pos – **P:** 0.81 **R:** 0.39 **F:** 0.52 **Tamil - Naïve Bayes**	**Without SentiWordNet** NB - **A:** 61.79% Maxent – **A:** 59.55% DT - **A:** 64.04% SVM – **A:** 71.91% **With SentiWordNet** NB - **A:** 66.17% Maxent – **A:** 64.04%	**Bengali** 2-class – **A:** 67.83% 3-class – **A:** 51.25% **Hindi** 2-class – **A:** 81.57% 3-class – **A:** 56.93% **Tamil** 2-class – **A:** 62.16% 3-class – **A:** 45.24%	**Bengali** **A:** 65.16% **F:** 0.644 **Hindi** **A:** 72.01% **F:** 0.714 **Tamil** **A:** 88.23% **F:** 0.802	**with Lyric Features SVM+NB -** **A:** 75.7% **with Audio Features SVM+GMM -** **A:** 91.2%

ISSN: 2233-7857 IJFGCN

	Neg – **P:** 0.76 **R:** 0.81 **F:** 0.78 Neu - **P:** 0.70 **R:** 0.77 **F:** 0.73 Pos – **P:** 0.74 **R:** 0.39 **F:** 0.51 **Tamil - C4.5 Decision Tree** Neg – **P:** 0.81 **R:** 0.84 **F:** 0.82 Neu - **P:** 0.71 **R:** 0.83 **F:** 0.77 Pos – **P:** 0.77 **R:** 0.37 **F:** 0.50	DT: **A:** 66.29% SVM – **A:** 75.96%			
Observation	**Variations of languages, punctuation marks, and spellings makes task of sentiment analysis more difficult in Indian Languages.**	More word coverage can result in better performance.	**Used 2-class and 3-class classification with features as Word-n-grams, Character-n-grams, surface and SentiWordNet features.**	RNN can be used to improve performance of classification but has problem of long term dependency.	**Lyrics features are generated using Doc2Vec and also considered audio features for classification.**

Table 1: Continued…

Reference	[43]	[26]	[28]	[16]	[10]
Language	**Gujarati**	Tamil	**Tamil**	Bengali	**Marathi**
Technique Used	**SVM**	TF-IDF, TF-IDF + Domain Specific Tags(DST), Tweet Weightage	**HMM, NB, RNN**	MNB,SVM(SMO)	**Not Specified**
Lexicon Type	**Microblogging**	Microblogging	**Contents from various websites**	Microblogging	**Movie Reviews**
Dataset Size	**Not Specified**	100 Tamil movies and around 7,000 tweets	**Not Specified**	SAIL-2015 Tweets Dataset	**Not Specified**

ISSN: 2233-7857 IJFGCN

Dataset	**Gujarati Tweets**	Tamil Movie Tweets	**Online Tamil Contents**	SAIL 2015 Tweets Dataset- Bengali	**Movie Reviews in Marathi**
Resource used	**POS Tagging**	NLTK, Tamil Dictionary	**Not Used**	**ITRANS, WEKA, SentiWordNet**	**Gate Processor's Dictionary- ANNIE**
Evaluation Measures	**SVM** **A:** 92%	**TF-IDF ranking** **A:** 29.87% **TF-IDF + DST** **A:** 35.64% **Tweet Weightage** **A:** 40.07%	**Long phrases:** **A:** 71.1% **Intra sentential Negation:** A: 73% **Inter sentential Negation:** **A:** 70.8%	**MNB** U: **A:** 41.6% U+Bi: **A:** 40.2% U+Senti: **A:** 43.6% U+Bi+Senti: **A:** 44.2% U+Bi+Tri+Senti: **A:** 44% **SVM (SMO)** U: **A:** 37.2% U+Bi: **A:** 36.6% U+Senti: **A:** 45% U+Bi+Senti: **A:** 41.4% U+Bi+Tri+Senti: **A:** 42.2%	Not Measured
Observation	**Used POS tagging for feature extraction and SVM for classification of sentiment.**	Tweet Weightage model has better accuracy than other and also need more improvements in linguistic models.	**RNN model can be used for improving sentiment analyzer tool.**	Correctly labeled dataset can help to improve performance of system.	**ANNIE (Gate Processor's Dictionary) is used to find polarity of word which is available in dictionary.**

Table 1: Continued…

Reference	[41]	[3]	[21]	[36]	[23]

ISSN: 2233-7857 IJFGCN

Language	**Marathi**	Hindi	**Hindi-English**	Bengali	**Malayalam**
Technique Used	**Corpus Based, Yandex Translator is used to translate to Marathi.**	Binary relevance approach and Label power set approach, NB, DT and SVM(SMO)	**Machine Learning Model MNB and Deep Learning classifier LSTM**	RNN-LSTM	**Lexicon Based, Naïve Bayes**
Lexicon Type	**Not Specified**	Product Reviews, Newspaper, blogs Collected from websites	User Comments from various websites	Microblogging	**Reviews, User Comments Collected from websites**
Dataset Size	**Not Specified**	5417 Reviews of 12 diff. domains, 2,250 positive, 635 negative, 2,241 neutral and 128 conflict instances of aspect categories	**3879 sentences split into 15% Negative, 50% neutral and 35% positive classes.**	SAIL-2015 Tweets Dataset	**136 Reviews, +ve – 11.76%, -ve – 65.44%, Neu – 11.76%**
Dataset	**Not Specified**	Product Reviews	**User Comments**	SAIL 2015 Tweets Dataset	**Malayalam User Reviews**
Resource used	**English SentiWordNet 3.0**	WEKA(MEKA), NB, DT, SMO	**Machine Learning and Deep Learning Classifier**	Bengali SentiWordNet	**Stop Words**
Evaluation Measures	Not Measured	**Electronics** NB: **A:** 50.95% DT: **A:** 54.48% SMO: **A:** 51.07% **Mobile Apps** NB: **A:** 46.78% DT: **A:** 47.95% SMO: **A:** 42.10% **Travels** NB: **A:** 56.06% DT: **A:** 65.20% SMO: **A:** 60.63% **Movies** NB: **A:** 87.78% DT: **A:** 91.62% SMO: **A:** 91.62%	**MNB+LSTM:** **A:** 70.8% **P:** 0.718 **R:** 0.612 **F:** 0.661	**RNN+LSTM** **A:** 55.27%	**Naive Bayes** **A:** 89.33% **P:** 88.88% **R:** 82.75% **F:** 85.7%

ISSN: 2233-7857 IJFGCN

Observation	**Performance of system depends on accuracy of Machine Translation Used.**	Considered 12 domains for aspect term detection Multi-label and Multi-class classification.	**Sentiment of sentence is predicted using combination of LSTM and MNB.**	Performance of system depends on size of training dataset, and number of correctly labelled dataset.	**Corpus polarity is calculated using positive and negative values of documents.**

Table 1: Continued…

Reference	[8]	[5]	[40]	[27]	[22]
Languages	Hindi	**Konkani**	Telugu	**English-Telugu**	English-Hindi
Technique Used	Adjectives represents Product Opinion and Noun Represents Feature	**NBB classifier**	Rule- Based Approach and Machine Learning based	**Lexicon based**	Baseline model, Lexicon-based model, Naïve Bayes model
Lexicon Type	Product Reviews	**Poems Written by various Authors**	Movie Reviews	**Microblogging**	Code-Mixed Corpus
Dataset Size	1000 Reviews of Mobile Phones	**50-poems Written by 22 Contemporary poets. And 10 randomly selected poems**	60 Telugu Movie Reviews, 783 Opinion words, 401 +ve,187 -ve, 195, Neutral words	**352 - Tweets English-Telugu**	Not Specified
Dataset	Mobile Phones Product Reviews	**Poems**	Telugu Movie Reviews	**Telugu Movie Tweets**	SAIL-2017
Resource used	Hindi POS Tagger(CDAC)	**Not Used**	POS Tagging, Telugu SentiWordNet	**Telugu-SentiWordNet, English Opinion Lexicon**	ITRANS, NLTK, POS Tagger,

ISSN: 2233-7857 IJFGCN

Evaluation Measures	Not Measured	**Naive Bayes** **50-Poems** **A:** 82.67% **P:** 71.15% **R:** 69.45% **F:** 60.15% **10-Poems** **A:** 70%	783 opinion words from which 51% are positive opinion words, 24% negative and 25% are Normal words.	**Using SentiWordNet** **A:** 79.9%	**Baseline model** **A:** 64.25% **Lexicon-based** **A:** 75% **NB model** **A:** 79.69%
Observation	**Self-learning ontologies can be developed for better accuracy.**	**Improved lexical resources results in better performance of system.**	Size of training and testing dataset has impact on performance of the system.	**Variations in spelling, and incorrect punctuation makes impact on performance of system.**	In code switching syntactic rules plays important role on performance.

A: Accuracy **P:** Precision **R:** Recall **F:** F-Measure **Neg:** Negative **Neu:** Neutra**l Pos:** Positive **U:** Unigram **Bi:** Bigram **Tri:** Trigram **Senti:** SentiWordNet **RHMR:** Rating Based Hindi Movie Reviews**, SMRD:** Standard Movie Reviews Dataset

Table 2: Resources available for Indian Languages

Sr. No.	NLP Resources	Languages
1	IndoWordNet- A WordNet of Indian Languages	हिन्दी (Hindi),English, অসমীয়া (Assamese),বাংলা (Bengali), बोडो (Bodo), ગુજરાતી (Gujarati), ಕನ್ನಡ (Kannada), کشر (Kashmiri), कोंकणी(Konkani), മലയാളം(Malayalam), मनिपुरि (Manipuri), मराठी (Marathi), नेपाली (Nepali), संस्कृतम् (Sanskrit), தமிழ் (Tamil), తెలుగు} (Telugu), ਪੰਜਾਬੀ (Punjabi), اردو (Urdu), ଓଡିଆ (odiya)
2	SentiWordNet	Hindi, Bengali, Telugu, Tamil, Urdu, Kannada, Oriya, Malayalam, Punjabi, Nepali, and Konkani.
3	Morphological Analyzers	Tamil v-2.0, Hindi Ver-6.0.0, Bengali 2.6.2, Kannada v-2.4, Marathi v-1.2,Punjabi v-1.9, Urdu v-1.4,
4	CRF Chunker	Punjabi CRF Chunker 1.4,Marathi CRF Chunker 2.4, Kannada CRF Chunker 1.3, Hindi CRF Chunker 1.5, Bengali CRF Chunker 1.2, CRF Chunker Tamil, CRF Chunker Urdu,
5	CRF POS Tagger	CRF POS Tagger 1.5 Tamil, CRF POS Tagger 1.5 Kannada, CRF POS Tagger 2.2 Hindi, CRF POS Tagger 1.2 Bengali, Punjabi CRF POS Tagger 1.4, CRF POS Tagger 1.4 for Urdu,
6	Shallow Parser	Tamil, Hindi, Marathi, Telugu, Punjabi, Urdu, Bengali

ISSN: 2233-7857 IJFGCN

Table 3 gives some Bilingual and Multilingual dictionaries available for Indian languages [1].

Table 3: Dictionaries for Indian Languages

Sr. No.	List of Dictionary	Languages
1	Multilingual dictionary	Tamil - Telugu, Hindi - Telugu, Urdu - Hindi, Tamil - Hindi, Bengali - Hindi, Panjabi - Hindi, Marathi - Hindi, Malayalam - Tamil, Kannada - Hindi
2	Bilingual Dictionary	Urdu to Hindi, Marathi to Hindi, Punjabi to Hindi, Malayalam to Tamil, Kannada to Hindi, Bangla to Hindi, Sanskrit–Hindi Apte Dictionary, Tamil-English

Table 4 shows different Corpus available online at Indian Language Technology Proliferation and Deployment Centre [1]. And *Table 5* shows some Lexical Resources available online at Indian Language Technology Proliferation and Deployment Centre [1].

Table 4: Different Corpus available for Indian Languages

Sr. No.	List of Text Corpus	Languages
1	Stop Words	Telugu, Tamil, Punjabi, Odia, Marathi, Hindi, Gujarati, Bengali, Assemese
2	Synset	Telugu, Tamil, Sanskrit, Nepali, Marathi, Manipuri, Malayalam, Kashmiri, Kannanda, Bodo, Bengali, Assamese, Hindi, Urdu, Punjabi, Oriya, Kokani, Gujarati,
3	Hindi - IL* Parallel Chunked Text Corpus ILCI-II	Hindi - Tamil, Punjabi, Nepali, Marathi, Malayalam, Konkani, Kannada, Assamese, Urdu, Bangla, Bodo, English, Gujarati, Odia
4	Author wise Marathi Language Text Corpus	Dataset–I is a collection of articles on category comedy by 5 different authors. A number of words by each author is ranging from minimum 7006 and maximum 10,411 words. Dataset–II is composed of articles of the mixed category. In total 10 different authors with minimum 26874 and maximum 33722 words.
5	Gujarati News Corpus - SRIMCA	The corpus covers news articles from reputed newspapers published in Gujarati. Currently it contains 156, 210, 101 and 50 news articles in the domain of business, crime, politics and sports.
6	Treebank Data_IIITH	Malayalam, Kannada, Marathi, Hindi, Bengali
7	Monolingual Text Corpus ILCI-II (General Domain)	Hindi, Urdu, Tamil, Punjabi, Odia, Nepali, Marathi, Malayalam, Konkani, Kannada, Gujarati, English, Bodo, Bangla, Assamese
8	Text Corpus - AnglaMT	Malayalam English General, Malayalam Monolingual Tourism, Malayalam Monolingual Health, Punjabi Monolingual, English Monolingual General, English Monolingual Health, English Monolingual Tourism

ISSN: 2233-7857 IJFGCN

9	General Text Corpus	Hindi – Marathi, Hindi – Punjabi, Telugu – Tamil , Hindi Bengali, Bengali To Hindi, Hindi To Kannada, Hindi To Tamil, Hindi To Telugu, Hindi To Urdu , Kannada To Hindi, Marathi To Hindi, Punjabi Hindi , Tamil Hindi, Tamil Telugu, Telugu Hindi, Urdu Hindi, Nepali, Manipuri, Bodo, Assamese
10	Tourism Text Corpus-ILCI** (25k Sentences)	Hindi-Urdu, Hindi-Telugu, Hindi-Tamil, Hindi-Punjabi, Hindi-Marathi , Hindi-Malayalam, Hindi-Konkani, Hindi-Gujarati, Hindi-English, Hindi-Bangla ,
11	Health Text Corpus-ILCI (25k Sentences)	Hindi-Telugu, Hindi-Punjabi, Hindi-Tamil, Hindi-Marathi, Hindi-Malayalam, Hindi-Konkani, Hindi-Gujarati, Hindi-Bangla, Hindi-English, Hindi-Urdu,

****ILCI**- Indian Languages Corpora Initiative (ILCI) project initiated by the MeitY, Govt. of India, Jawaharlal Nehru University, New Delhi had collected corpus in Hindi as source language and translated it in ILs as the target language [1].

Table 5: Lexical Resources for Indian Languages

Sr. No.	List of Lexicon	Languages
1	EILMT Health Lexicon	English-Urdu , English-Tamil , English-Odia , English-Marathi, English-Hindi, English-Gujarati, English-Bodo, English-Bangla
2	EILMT* Agriculture Lexicon	English-Urdu ,English-Tamil,English-Odia, English-Marathi, English-Hindi, English-Gujarati, English-Bodo, English-Bangla
3	EILMT Tourism Lexicon	English-Tamil, English-Urdu, English-Odia, English-Marathi, English-Hindi, English-Bangla
4	Transliteration Lexicon	Tamil-Hindi, Tamil-English

*English to Indian Language Machine Translation (EILMT)

5. Conclusions:

This review paper covers various sentiment classification techniques, lexical resources available for Indian languages to perform sentiment analysis. As Indian language content are increasing rapidly over the internet, so it is necessary to build a resources for low resourced Indian languages. There are many challenges that we have to overcome while developing the resources. But there are some languages which are highly resourced as Hindi, Bengali, Tamil, and Telugu which helped them to perform sentiment analysis on these Indian Languages. As the interest in Indian languages has been increasing rather than English, but still some Indian languages are under resourced which has biggest concern for the researcher and opportunity to develop quality resources for these languages.

There are many resource collected from microblogs, blogs, newspaper, websites and author's corpus to perform Sentiment Analysis on Indian Languages. And Sentiment classification techniques are used such as Machine learning, lexicon based and Hybrid approaches. In Machine learning algorithms used are SVM, NB, Decision Tree and various lexicon are also used such as SentiWordNet, WordNet, domain specific

ISSN: 2233-7857 IJFGCN

corpus and Stop words. As many Indian languages are unexplored and these languages should be considered for future work to perform Sentiment Analysis. And we are considering Marathi language as one of the under resourced language for building resources and performing sentiment analysis.

REFERENCES

1. (2019). Retrieved 05 16, 2020, from Technology Development for Indian Languages (TDIL): https://tdil-dc.in/index.php?option=com_download&task=fsearch&Itemid=547&lang=en
2. A. Das, & S. Bandyopadhyay. (2010). SentiWordNet for Indian Languages. *The 8th Workshop on Asian Language Resources* (pp. 56-63). Beijing, China: Coling 2010. Retrieved 05 16, 2020, from https://www.aclweb.org/anthology/W10-3208.pdf
3. Akhtar M.S, Ekbal A, & Bhattacharyya P. (2018). Aspect Based Sentiment Analysis: Category Detection and Sentiment Classification for Hindi. *Computational Linguistics and Intelligent Text Processing. CICLing 2016. 9624*, pp. 246-257. Springer, Cham. doi:https://doi.org/10.1007/978-3-319-75487-1_19
4. Anagha M,, Raveena R, K., Sreetha k, & P C Reghu Raj. (2015). Fuzzy Logic Based Hybrid Approach for Sentiment Analysis of Malayalam Movie Reviews. *IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)* (pp. 1-4). Kozhikode: IEEE.
5. Annie Rajan, & Ambuja Salgaonkar. (2020). Sentiment Analysis for Konkani Language: Konkani Poetry, a Case Study. *ICT Systems and Sustainability. Advances in Intelligent Systems and Computing* (pp. 321-329). Springer, Singapore.
6. Arunselvan S J, Anand kumar M, & Soman K P. (2015). Sentiment Analysis of Tamil Movie Reviews via Feature Frequency Count. *International Journal of Applied Engineering Research, 10*(20), 17934-17939.
7. Ayush Kumar, Sarah Kohail, Asif Ekbal, & Chris Biemann. (2015). IIT-TUDA: System for Sentiment Analysis in Indian Languages Using Lexical Acquisition. *Springer, Cham*, 684-693.
8. Bandari, S., & Bulusu, V. (2020). Survey on Ontology-Based Sentiment Analysis of Customer Reviews for Products and Services. *Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing* (pp. 91-101). Springer, Singapore.
9. Baxi, A. (2018, March 29). *Tech Startups, Take Note: More Indians Access The Internet In Their Native Language Than In English*. Retrieved from https://www.forbes.com/: https://www.forbes.com/sites/baxiabhishek/2018/03/29/more-indians-access-the-internet-in-their-native-language-than-in-english/#4b2d9d1a4a03
10. Chitra V. Chaudhari, Ashwini, Rashmi, & Komal. (2017). Sentiment Analysis in Marathi using Marathi Wordnet. *Imperial Journal of Interdisciplinary Research (IJIR), 3*(4), 1253-1256.
11. Diwanji, S. (2020, April 1). *Digital population across India*. Retrieved from www.Statista.com: statista.com/statistics/309866/india-digital-population
12. Harika, Akkireddy, Suryakant, & Radhikai. (2016). Multimodal Sentiment Analysis of Telugu Songs. *4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016)* (pp. 48-52). New York City, USA: IJCAI 2016.
13. *IndoWordNet : A wordnet of Indian languages*. (2020, 05 16). Retrieved 05 16, 2020, from http://www.cfilt.iitb.ac.in/: http://www.cfilt.iitb.ac.in/indowordnet/
14. Jain, S., & Shashank, B. (2015). Cross-Lingual Sentiment Analysis using modified BRAE. *Conference on Empirical Methods in Natural Language Processing* (pp. 159-168). Lisbon, Portugal: Association for Computational Linguistics.
15. Jha, V., Manjunath N, Shenoy, P., Venugopal K R, & L M Patnaik. (2015). HOMS: Hindi Opinion Mining System. *IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)* (pp. 366-371). Kolkata, India: IEEE.

ISSN: 2233-7857 IJFGCN

16. Kamal Sarkar, & Mandira Bhowmick. (2017). Sentiment Polarity Detection in Bengali Tweets Using Multinomial Naïve Bayes and Support Vector Machines. *2017 IEEE Calcutta Conference (CALCON)* (pp. 31-35). kolkata, India: IEEE.
17. *Languages of India*. (2020, April 24). Retrieved from Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Languages_of_India
18. Liu, B. (2017). Many Facets of Sentiment Analysis. In E. Cambria, D. Das, S. Bandyopadhyay, & A. Feraco (Eds.), *A Practical Guide to Sentiment Analysis. Socio-Affective Computing* (Vol. 5, pp. 11-39). Springer, Cham.
19. M. A. Ansari, & S. Govilkar. (2016). Sentiment Analysis of Transliterated Hindi and Marathi Script. *Sixth International Conference on Computational Intelligence and Information* (pp. 142-149). Cochin, India: McGraw-Hill Education.
20. M.Alshari, A. Azman, Doraisamy, Mustapha, & Alkeshr. (2018). Effective Method for Sentiment Lexical Dictionary Enrichment based on Word2Vec for Sentiment Analysis. *2018 Fourth International Conference on Information Retrieval and Knowledge Management* (pp. 177-181). IEEE.
21. Madan Gopal Jhanwar, & Arpita Das. (2018). An Ensemble Model for Sentiment Analysis of Hindi-English Code-Mixed Data. *Workshop on Humanizing AI (HAI)* . Stockholm, Sweden: IJCAI.
22. Mahata S.K., Makhija S., Agnihotri A., & Das D. (2020). Analyzing Code-Switching Rules for English–Hindi Code-Mixed Text. *Emerging Technology in Modelling and Graphics. Advances in Intelligent Systems and Computing. 937*, pp. 137-145. Springer, Singapore. doi:https://doi.org/10.1007/978-981-13-7403-6_14
23. Mathews D.M., & Abraham S. (2019). Twitter Data Sentiment Analysis on a Malayalam Dataset Using Rule-Based Approach. In S. N., P. L., N. H., H. P., & N. N. (Ed.), *Emerging Research in Computing, Information, Communication and Applications. 906*, pp. 407-415. Springer, Singapore.
24. Md Shad Akhtar, Asif Ekbal, & P. Bhattacharyya. (2016). Aspect based Sentiment Analysis in Hindi: Resource Creation and Evaluation. *Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 2703-2709). Portorož, Slovenia: European Language Resources Association (ELRA). Retrieved from https://www.aclweb.org/anthology/L16-1429
25. Mohammed Arshad Ansari, & Prof. Sharvari Govilkar. (2016). Sentiment Analysis of Transliterated Hindi and Marathi Script. *Sixth International Conference on Computational Intelligence and Information* (pp. 142-149). Cochin, India: McGraw-Hill Education.
26. Nadana Ravishankar, & Shriram. (2017). Corpus based Sentiment Classification of Tamil movie tweets using Syntactic patterns. *IIOAB Journal, 7*(7), 1-3.
27. Padmaja S., Fatima S., Bandu S., Nikitha M., & Prathyusha K. (2020). Sentiment Extraction from Bilingual Code Mixed Social Media Text. *Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing. 1079*, pp. 707-714. Springer, Singapore. doi:https://doi.org/10.1007/978-981-15-1097-7_59
28. Padmamala, R., & Prema, V. (2017). Sentiment Analysis of Online Tamil Contents using Recursive Neural Network Models Approach for Tamil Language. *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)* (pp. 28-31). Chennai, T.N., India: IEEE.
29. Pandey, P., & Sharvari, G. (2015). A Framework for Sentiment Analysis in Hindi using HSWN. *International Journal of Computer Applications, 119*(19), 23-26.
30. Parul Sharma, & Teng-Sheng Moh. (2016). Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter. *IEEE International Conference on Big Data (Big Data)* (pp. 1966-1971). Washington, DC: IEEE.
31. Prof. Sumitra Pundlik, Prachi Kasbekar, & Gajanan Gaikwad. (2016). Multiclass Classification and Class based Sentiment Analysis for Hindi Language. *2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 512-518). Jaipur, India: IEEE.

ISSN: 2233-7857 IJFGCN

32. S. Deshmukh, NILEEMA, SURABHI, & JASON. (2017). SENTIMENT ANALYSIS OF MARATHI LANGUAGE. *International Journal of Research Publications in Engineering and Technology [IJRPET], 3*(6), 93-97.
33. S. Phani, S. Lahiri, & A. Biswas. (2016). Sentiment Analysis of Tweets in Three Indian Languages. *6th Workshop on South and Southeast Asian Natural Language Processing*, (pp. 93–102). Osaka, Japan.
34. S. Pundlik, P. Kasbekar, & G. Gaikwad. (2016). Multiclass Classification and Class based Sentiment Analysis for Hindi Language. *2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 512-518). Jaipur, India: IEEE.
35. S. S. Prasad, J. Kumar, D. K. Prabhakar, & S. Tripathi. (2016). Sentiment mining: An approach for Bengali and Tamil tweets. *2016 Ninth International Conference on Contemporary Computing (IC3* (pp. 1-4). Noida: IEEE.
36. Sarkar, K. (2019). Sentiment Polarity Detection in Bengali Tweets Using LSTM Recurrent Neural Networks. *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)* (pp. 1-6). Gangtok, India: IEEE.
37. Shanta Phani, Shibamouli Lahiri, & Arindam Biswas. (2016). Sentiment Analysis of Tweets in Three Indian Languages. *6th Workshop on South and Southeast Asian Natural Language Processing*, (pp. 93–102). Osaka, Japan.
38. Shriya Se, R. Vinayakumar, M. Anand Kumar, & K. P. Soman. (2016). Predicting the Sentimental Reviews in Tamil Movie using Machine Learning Algorithms. *Indian Journal of Science and Technology, 9*(45), 1-5.
39. Shriya Seshadri, Anand Kumar, & Soman Kotti P. (2016). ANALYZING SENTIMENT IN INDIAN LANGUAGES MICRO TEXT USING RECURRENT NEURAL NETWORK. *THE IIOAB Journal, 7*(1), 313-318.
40. Srinivasu Badugu. (2020). Telugu Movie Review Sentiment Analysis Using Natural Language Processing Approach. *Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing. 1079*, pp. 685-695. Springer, Singapore. doi:https://doi.org/10.1007/978-981-15-1097-7_57
41. Sujata Deshmukh, Nileema, Surabhi, & Jason. (2017). SENTIMENT ANALYSIS OF MARATHI LANGUAGE. *International Journal of Research Publications in Engineering and Technology [IJRPET], 3*(6), 93-97.
42. Sujata Rani, & Parteek Kumar. (2018). A journey of Indian languages over sentiment analysis: a systematic review. *Springer*, 1415-1462.
43. Vrunda C. Joshi, & Dr. Vipul M. Vekariya. (2017). An Approach to Sentiment Analysis on Gujarati Tweets. *Advances in Computational Sciences and Technology, 10*(5), 1487-1493.
44. Yakshi Sharma, Veenu Mangat, & Mandeep Kaur. (2015). A Practical Approach to Sentiment Analysis of Hindi Tweets. *1st International Conference on Next Generation Computing Technologies (NGCT-2015)* (pp. 677-680). Dehradun, India: IEEE.

ISSN: 2233-7857 IJFGCN