

Machine Learning Techniques for Sentiment Analysis of Indian Languages

Gazi Imtiyaz Ahmad, Jimmy Singla

Abstract— Sentiment Analysis is the domain of automatically understanding the emotions, feelings, opinions in a textual data. It is a way of understating how a product, brand, service, idea or an event is viewed by common people, customers and stakeholders. Sentiment Analysis Systems are used by politicians, business leaders, developers and researchers to infer useful information as per their specific needs. It is used in business decision making process to value the views of the customers. Sentiment analysis has become a hot topic of scientific and market research in the field of natural Language Processing. India is a large populated country and the number of Internet users is also huge. Most people share their experience in English. However, during the last decade, due to the accessibility of Internet and evolution in language modelling people express their views in their own native Indian language. With the increase in Indian language text, researchers find it quite fascinating to infer valuable information from this unstructured text data. A number of machine learning techniques have been applied on this textual data set. Basic concepts of Sentiment analysis shall be discussed with focus on Indian language text in this paper. Due to on availability of rich lexicon resources for unsupervised learning techniques and better evaluation measures for the Supervised learning techniques, the later become the first choice for researchers in the field of Natural Language Processing. A comparative analysis shall be made for various supervised machine learning techniques in the context of Indian languages.

Index Terms: Machine Learning, Natural Language Processing, Sentiment Analysis, Supervised Learning

I. INTRODUCTION

With the popularity of social media, microblogging and ecommerce sites and increased Internet population, people indicate their views, opinions, feelings, emotions, and reviews in a free form and without any external influence. The accumulation of unstructured textual data on the web encourages researchers to deduce useful information from it which steered further research in NLP area known as “Opinion mining and Sentiment analysis”. Sentiment analysis can be defined as the automatic mining of feelings, reviews, opinions, emotions for a product, service, entity, idea from textual data through Natural language processing. It involves classification of opinions into three categories. The first being as “positive”, the second category as “negative” and third category as “neutral” [1]. Opinion mining is the analysis of user reviews, feelings, emotions posted on microblogging and e-commerce sites about a person, product, process, idea, or service. Today we use Natural Language processing techniques to extract polarity in a textual data. Sentiment

analysis process starts from the data collection to the evaluation process and or visualization. Advancements in Artificial Intelligence and machine learning techniques have brought sentiment analysis into the forefront of cutting-edge algorithms.

India is a multilingual culturally and linguistically diverse country having a population of more than one billion [3]. People of different cultures ad regions within the country use different languages to communicate and post their opinions on social platforms and blogging websites. In its ICUBE 2018 report, market research agency Kantar IMRB, has reported that due to easier access to Internet, because of low-priced data plans including government funded publicity workshops and bandwidth availability, the Internet users in India is estimated at 566 million as of December, 2018 and is expected to reach 627 million (about 45% of the population) in 2019. According to Internet Search giant Google, the internet users who communicate in Indian languages are predicted to reach 536 million over next four years. Also the social media users in India are around 326 million. The online communication of Indian users on Social media and networking sites has increased tremendously. The accumulation of large volume of Indian language data on the Internet has opened ways for researchers and business leaders to interpret this data and transform it into more useful information by way of “Opinion mining and Sentiment analysis”. Although Sentiment analysis of textual data is performed at document, sentence and aspect level, most research work has been carried out by researchers at sentence level. This is because Tweets and opinions on movie/product reviews expressed by the users are in short passages or segments and researchers find it quite useful for their research work.

During the past decade the research in sentiment analysis of Indian language text has got momentum. However, due to inaccessibility of labeled datasets for Indian languages, the task of sentiment analysis becomes more arduous. Sentiment analysis is a challenging task as People express their opinions in complex ways. Identifying sentiments in such type of environment needs thorough contextual understanding. Also most of the sentiment analysis happens in categorical framework, therefore a sentiment is classified and assigned to a certain category. Availability of pre-trained models in data science languages makes it difficult in deciding the process of training the machine learning model that can be implemented for specific work scenario. Also, personal analysis by an individual on a particular event, product or service contrasts with others opinion and is mostly influenced by socio-cultural parameters including individual’s personal experience.

Revised Version Manuscript Received on 16 September, 2019.

Gazi Imtiyaz Ahmad, Research Scholar, Department of Computer Applications LPU Punjab.

Dr. Jimmy Singla, Associate Professor, Department of CSE, LPU Punjab. jimmy.21733@lpu.co.in

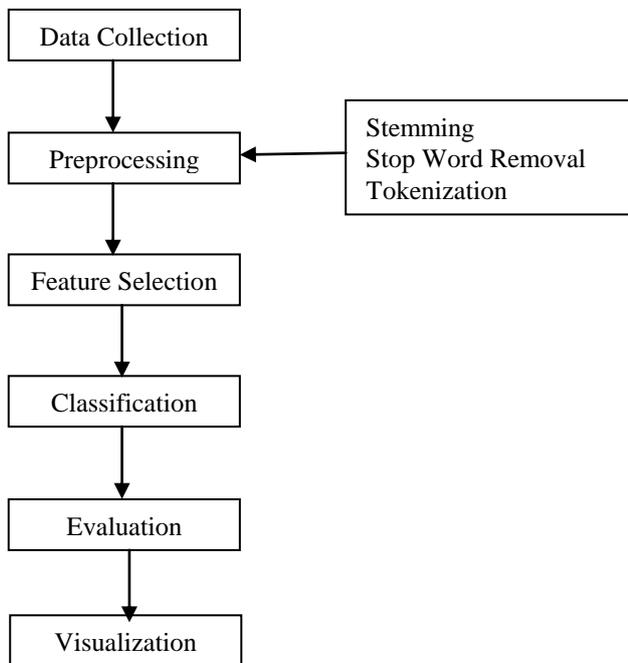


Fig: 1 Sentiment Analysis Process

1. Data Collection:

User generated textual data can be in the form of weblogs, discussion forums, news, reviews and on social networking sites such as Twitter, Facebook, Instagram, Google+ etc. Labelled datasets have already been created especially for Twitter posts and product reviews. Data can also be collected using APIs provided by many platforms and forums.

2. Preprocessing

Preprocessing forms the fundamental step in sentiment analysis system for text mining to prepare data for classification [4]. This phase include text cleaning, feature extraction and selection. A number of techniques are applied to textual data for text noise reduction, minimizing dimensionality which helps in effective classification. These include removal of unnecessary and unimportant entries/elements/items from the text such as URLs, Hashtags, line breaks, vowels, replacement of negative constructs with “negation”, interchanging slang words with its conventional interpretation, stemming to decrease entropy, and filtration of stop words. Sentences are split into words, phrases, symbols and other meaningful tokens by removing punctuation marks. POS tagging is done for recognition of different parts of speech in the text. The text is finally converted into lower case. Feature extraction and feature selection are two important steps in preprocessing. Commonly used feature evaluation metrics include information gain, term frequency, and inverse term frequency, ch-square, expected cross entropy, Odds Ration and Gini index [5].

3. Classification

Classification is the process of identification of set of categories/classes for a new observation on the basis of training set. Classification is associated with the supervised learning where a classifier understands from a training set of data and predicts how a new input is related to a particular class. A number of techniques exist for Sentiment classification. However, the two most frequently used techniques include lexicon based implementation and operating machine learning algorithms. Apart from these two

we also have blended techniques that include amalgamation of the above mentioned two approaches in addition to deep learning techniques which is a novel approach in Natural language processing and has got researcher’s attention in the recent past.

a. Lexicon based approach

Lexicon based technique is hinged with the assumption that identification of word polarity or text segment is influenced by adjacent words that compose it [5]. A sentiment lexicon is created through manual approach by assigning polarity to the words by humans which can be a time consuming process. A dictionary-based approach adds words and their synonyms and antonyms iteratively in order to create a lexicon. Corpus based approach depends on syntactic patterns in a large data set [6]. The word in the document to be analyzed is matched with the lexicon and polarities are accordingly assigned. The individual word polarities in the text is calculated and then aggregated to determine the semantic orientation of the entire text.

b. Machine Learning approach

Machine learning forms the most robust and innovative platforms for learning inter-relations between data for development of organization to gain a new level of understanding. Machine learning is the subcomponent of Artificial Intelligence area that takes data as input and the model learns from given patterns and data features instead of using traditional ways of programming. Machine learning operates with diverse algorithms that studies interrelations and given patterns in data in order to achieve optimal prediction and accuracy. Machine learning is categorized into three categories which “supervised”, “unsupervised” and “semi-supervised learning”.

i. Supervised Learning:

Supervised learning has well-known set of data and an understating of classification of that data. A training data set and a test data set is used for classification. The training data set helps the classifier to lean and model relationship and dependencies between the two data sets.

The data has labelled features that define the meaning of data. Supervised learning forecast the output values of new input data based on the relationships it learned from the training data. Weather forecasting is a good example of supervised learning. Supervised machine learning techniques such as Naïve Bayes, Support Vector Machines have achieved notable success in text classification [41].

ii. Unsupervised learning:

Unsupervised learning is commonly used where large volume of unlabeled data is required. Unsupervised machine learning algorithms do not require labelled data and can be processed by clustering methods [39]. These classification techniques are used pattern detection. For example, studying relationship between symptoms of various diseases by analyzing humongous volumes of medical data available.

iii. Semi-supervised learning:

Semi-supervised learning used the approach of both supervised and unsupervised machine learning methods. It works with minimal annotated data in amalgamation with a substantial volume of unlabeled data. The algorithm



operates with annotated data including the data for which labels are unknown. The goal is to classify unknown and unlabeled data using the information from the labelled data set.

c. Hybrid approach:

Hybrid or blended techniques includes mixture of both machine learning algorithms and lexicon based techniques for classification [7]. The principal advantage offered by hybrid techniques is that we can achieve combined accuracy of both lexicon techniques and machine learning based classification. Combining the two, the algorithm can improve evaluation measures of accuracy, precision and recall

Supervised Machine Learning Algorithms

Decision Tree

Decision tree uses tree structure to build classification models. It uses hierarchical division of data where internal nodes represent an attribute condition with branches constituting the outcome. The class labels are characterized by leaf node.[40]. In the classification process, the impact of attributes decrease from root to leaf node. Attribute impact is identified using the information gain concept. Rules are learned using the training data set one at a time. A new learned rule eliminate training tuples. This process continuous on training data set till a termination condition is reached. The tree construction is a recursive divide-and-conquer approach.

Naive Bayes

Based on Bayes theorem, Naïve Bayes classifier is a supervised method based on the theory of probabilities that works on the principle that every feature in a dataset is equally important and independent. The algorithm works quite effectively on relatively substantial data sets demanding minimal processing power. Naïve Bayes is a simple algorithm with good result achievements. One of the disadvantages of the algorithm is that it is reasonably unsuccessful in predicting correct result for a given attribute when its conditional property is zero.

Support Vector Machine

A supervised classifier, Support Vector Machine separates data into two categories on an N-dimensional hyper-plane. The dimensions of the hyper-plane depends on the number of features chosen. Hyper-planes are decision boundaries which classify the data points. The data points on either side of a hyper-plane are attributed to different classes. A hyper-plane is two dimensional if the number of input features are three. Support vectors represent data points and influence the position and orientation of hyper-plane. The objective of SVM classifier is to augment the separating hyper plane between data points.

K-Nearest Neighbors

K Nearest Neighbors is one of the simple and most commonly used supervised learning classification algorithm. Being a non parametric and slow learning method, it studies the given training data and classifies a new input data rooted on similarity measures with respect to learned data. The classification of text depends on the surrounding neighbors and the sample is allocated to that class which is most frequent among its K-nearest neighbors. To obtain the neighbors, a distance function is used. However, for categorical data, the “Hamming distance” function is used.

Logistic Regression

One of the principle machine learning method employed for binary classification is the Logistic Regression technique. Using logistic functions and probability estimation, logistic regression measures the relationship between a dependent variable and one or more independent variables. The probabilities are converted into binary values of 0 and 1 for predicting the outcome of input variable. The conversion of probabilities into binary values is done by employing a logistic or sigmoid function and threshold classifier. The algorithm performs better when attributes which are unrelated to the output or which are corresponding to one other are eliminated. Thus feature selection is a substantial component affecting the performance of machine learning classifier. The algorithm is quite efficient and effortless to execute and train.

4. Evaluation

Four matrices are commonly used to evaluate the performance of a classifier. These are accuracy, precision, recall and F-measure. These four metrics are defined in terms of True Positive (tp), False Positive (fp), True Negative (tn) and False Negative (fn) rates. The other average performance parameters like macro and micro and weighted F1-values are also used.

Accuracy measures the corrected classified instances and is defined as

$$\text{Accuracy} = \frac{\text{Number of correctly classified instances}}{\text{Total number of instances}}$$

Precision measures the exactness of a classifier

$$\text{Precision} = \frac{\text{True Positive (tp)}}{\text{True Positive (tp)+ False Positive (f p)}}$$

Recall measures the completeness of a classifier.

$$\text{Recall} = \frac{\text{True Positive (tp)}}{\text{True Positive (tp) + False Negative (fn)}}$$

The single metric of combined precision and recall is called F-measure

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$$

5. Visualization

Well known techniques such as graphs, histograms, confusion matrices are used for visualization of evaluation process. World-cloud, interactive maps, Sparkline-style plots are also common.

Preprocessing of data is an important step in sentiment analysis. The data preprocessing eliminates uncertainty from the unorganized data and produces clean data for classification. This can be done by a number of ways viz. stemming, tokenization, POS tagging, stop word removal etc. The approaches to sentiment classification include machine



learning as well as lexicon-based techniques. Employing Machine Learning techniques usually result in greater accuracy while as semantic orientation offers superior generality [3]. Recent research reveals that supervised classifiers outperform unsupervised classifiers in classification accuracy. However, the supervised learning algorithms need annotated training datasets, but the efficiency of an unsupervised approach hinges on the standard of the lexical resources [8].

II. RELATED WORK

A lot of work has been done in the area of NLP of sentiment analysis and opinion mining. Different approaches have been used for resource poor Indian languages as well. During the early years, lexicon methods were the preferred choice of researchers but because of unavailability of lexicon resources and annotated data sets, there has been a transition to the machine learning approach.

The authors in [8] provides a systematic survey on sentiment analysis for Indian language families. Sentiment analysis based on methods, sentiment levels, diverse domains and classes has been presented in a detailed manner. Earlier similar attempts were made in [9][10][11] but were limited to a fixed number of languages. The research in [12] presents a comparative study of opinion mining and sentiment analysis techniques; however it's restricted only to Indian language domain.

Machine Learning Techniques of sentiment analysis for Indian languages:

English language dominates the research work in the field of Natural Language Processing owing to its rich lexicon and availability of annotated resources. However, research work in the field of Natural Language Processing in context of Indian languages has shown positive signs from the past decade

The first attempt in sentiment analysis for Indian languages was started by [13] for Hindi language. The authors used unsupervised method and achieved an accuracy of 78.14%. The authors in [14] used decision tree algorithm for sentiment classification of Hindi tweets under constrained and unconstrained environment and achieved an accuracy of 40.47% and 31.26% respectively. Authors in [15] used SVM and Decision Tree for Hindi tweet in a constrained environment using TF-IDF scores of unigrams and tweet specific features and achieved a test accuracy of 43%. Authors in [16] presented an approach for Classification of Bangla text documents based on inverse class frequency.

The feature engineering method that was employed was based on TF-IDF-ICF coupled with Naïve Bayes Multinomial classification approach. The classification accuracy obtained was 98.87%. In [17], authors present sentiment classification framework using aspect-based opinion mining that was evaluated using dataset containing Indian movie reviews. The work presents a comparative study of feature extraction methods, as well as machine-learning classifiers. The results depict that SVM performs comparatively better than classifiers. Authors in [18] proposed a Neural Network-Based Architecture for Sentiment Analysis in Indian Languages. The technique evaluates 39 diverse combined frameworks of neural network layers and found that CNN (Convolutional Neural Network) obtained best accuracy of 77.63%.

In [19], authors proposed Sentiment identification of reviews posted by customers in Roman Urdu about automobiles. Their work reports that Multinomial Naïve Bayes performs better than other classification algorithms and obtains higher data values on performance parameters such as accuracy, precision, recall and F-measure. Authors in [20] proposed Sentiment Analysis of Movie Reviews in Hindi Language using Machine Learning and the results showed that the SVM classifier performed better than other machine learning techniques.

S. No.	Author	Language	Data Set used	Features used	Machine learning algorithm	Performance measure			
						Accuracy	Precision	Recall	F-measure
1	Prasad et al [22]	Hindi	Tweets	information gain ratio	DT	40.47%	0.822	0.815	0.804
2	Venugopalan et. al [23]	Hindi	Tweets	Unigram, tweet specific features	SVM	42.83%	-	-	-
3	Kumar et al [24]	Hindi, Bengali	Tweets	Character and Word,	SVM	47.96% (Hindi) 42.00 % (Bengali)	-	-	-
4	Sarkar & Chakraborty [24]	Hindi, Bengali	Tweets	unigrams, bigrams and trigrams	MNB	50.75 % (Hindi) 41.20 % Bengali)	-	-	-



					SVM	45.11 % (Hindi) 40.89 % Bengali)	-	-	-
5	Se S et. al [26]	Hindi							
6	Jha et al [27]	Hindi	Movie review	Unigrams, Best words and Best words + bigram chi square	NB	80.00%	0.834	0.77	0.80
7	Phani et. al [28]	Hindi Bengali Tamil	Tweets (SAIL Dataset 2015)	Word n-grams Character n-gram, Surface feature,	MNB, DT, LR, RF, SVM	56.96% (Hindi)			
						51.25% (Bengali)			
						45.24% (Tamil)			
8	Sharma & Moh [29]	Hindi	Tweets		NB, SVM	62.1% (NB) 78.4% (SVM)	0.71 0.75	0.61 0.78	0.65 0.76
9	Akhtar et. al [30]	Hindi	Movie reviews, Online Newspap er articles, blogs, product and service reviews	POS, Word Bigrams:	SVM	54.05%	0.62	0.30	0.41
10.	Das & Bandyopadhyay [31]	Bengali	Bengali News text	POS, chunk, functional word	SVM		0.70	0.63	0.66
11	Ghosal et. al [32]	Bengali	Online Newspap er articles on horescope	Unigram, bigram , trigram	NB	93.05%	0.93	0.93	0.93
					SVM	98.7%	0.98	0.98	0.98
					DT	95.16%	0.95	0.95	0.95
					RF	98.2%	0.98	0.98	0.98
12	Sharmista & Ramaswami [33]	Tamil	Product reviews	POS	RF	94.0%			
13	Nair et al [34]	Malayalam	Film reviews	n-gram	SVM	91%	0.806	0.951	0.863
14	Mukhtar and , Khan [35]	Urdu	Urdu blogs	First-neutral word	SVM	65.0%	0.67	0.65	0.64
				Second-neutr al word	DT	62.5%	0.66	0.62	0.60
				First-positive word Secon-positi ve word	k-NN	67.01%	0.67	0.67	0.67
15	Hegde & Padma [36]	Kannada	Mobile product reviews	Continuous and categorical aspect values	RF	72.0%	0.67	0.63	0.64

16	Kaur & Gupta [37]	Punjabi	Blogs and newspapers	n-gram	NB	75.31%	0.78	0.60	0.67
17	Mukku et al [38]	Telugu	Newspapers		NB, SVM, RF, LR, DT	60.13	0.56	0.59	0.56

III. RESULTS

The current research reports that supervised classifiers offer augmented performance when compared with unsupervised algorithms [21]. This is because data scientists have fewer techniques to build unsupervised models than for building supervised models. Also supervised classifiers are less complex and extremely precise than unsupervised models. However, the supervised methods requires large volume of labeled data which often become time consuming and expensive. As shown in table 1, in this study we have focused on only supervised learning techniques. It has been reported in the research that majority of the research work in the field of sentiment prediction of Indian language context is done on Hindi language (33%) followed by other languages such as Bengali (14%) and Tamil (10%) [8]. A good amount of research work has been done in Malayalam (9%), Urdu (7%) and Kannada (7%). Researchers have reported that Machine learning approach is used by most of the researchers (61%) followed by lexicon based approach (26%). Among the machine learning classifiers, SVM (28%), Naïve Bayes (25%) and Decision Tree (19%) are commonly used for polarity prediction. In addition most of the researchers use twitter (27%) data set for experimentation, however, movie (19%) and product (14%) reviews are also used by the researchers. The literature reports that sentiment analysis is mostly done at sentence level (72%) followed by document level analysis (18%). However, the recent research work has also been done at aspect level of sentiment analysis.

IV. CONCLUSION

Presently the first choice for performing sentiment analysis on text is employing machine learning methods because supervised learning algorithms perform better than lexicon based approach. Moreover, the accuracy rate of machine learning classifiers is better and these techniques are easy to use. Machine learning algorithms have their own benefits and limitations over one another, user preferences and context of the text play important role in choosing the algorithm. It has been observed that Naïve Bayes classifier as well SVM’s are the most frequently employed supervised machine learning algorithms. In future similar type of survey shall be conducted for code-mixed Indian languages

REFERENCES

- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1-135.
- Angiani, G., Ferrari, L., Fontanini, T., Fornacciar, P., Iotti, E., Magliani, F., & Manicardi, S. (2016). A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter. In *KDWeb*.
- Mahajan, P. (2009). Use of social networking in a linguistically and culturally rich India. *The International Information & Library Review*, 41(3), 129-136.

- Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
- Musto, C., Semeraro, G., & Polignano, M. (2014). A comparison of lexicon-based approaches for sentiment analysis of microblog posts. *Information Filtering and Retrieval*, 59.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.
- Jain, A. P., & Dandannavar, P. (2016). Application of machine learning techniques to sentiment analysis. In *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)* (pp. 628-632). IEEE.
- Rani, S., & Kumar, P. (2018). A journey of Indian languages over sentiment analysis: a systematic review. *Artificial Intelligence Review*, 1-48.
- Kaur, J., & Saini, J. R. (2014). 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(2), 53-60.
- Pandey, P., & Govilkar, S. (2015). A framework for sentiment analysis in Hindi using HSWN. *International Journal of Computer Applications*, 119(19).
- Govindan R, Haroon RP (2016) A survey on sentiment and emotion classification in Indo-Dravidian languages. *Imp J Interdiscip Res* 3(1):1040–1042.
- Haseena Rahmath, P., & Ahmad, T. (2014). Sentiment Analysis Techniques-A Comparative Study.
- Joshi, A., Balamurali, A. R., & Bhattacharyya, P. (2010). A fall-back strategy for sentiment analysis in hindi: a case study. *Proceedings of the 8th ICON*.
- Prasad, S. S., Kumar, J., Prabhakar, D. K., & Pal, S. (2015, December). Sentiment classification: an approach for Indian language tweets using decision tree. In *International Conference on Mining Intelligence and Knowledge Exploration* (pp. 656-663). Springer, Cham.
- Venugopalan, M., & Gupta, D. (2015, December). 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.
- Dhar, A., Dash, N. S., & Roy, K. (2018, February). Classification of Bangla Text Documents based on Inverse Class Frequency. In *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)* (pp. 1-6). IEEE.
- Trivedi, S. K., Dey, S., & Kumar, A. (2018). Capturing user sentiments for online Indian movie reviews: A comparative analysis of different machine-learning models. *The Electronic Library*, 36(4), 677-695.
- Bhargava, R., Arora, S., & Sharma, Y. (2018). Neural Network-Based Architecture for Sentiment Analysis in Indian Languages. *Journal of Intelligent Systems*.
- Khan, M., & Malik, K. (2018, April). Sentiment Classification of Customer’s Reviews About Automobiles in Roman Urdu. In *Future of Information and Communication Conference* (pp. 630-640). Springer, Cham.
- Nanda, C., Dua, M., & Nanda, G. (2018, April). Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning. In *2018 International Conference on Communication and Signal Processing (ICCSPP)* (pp. 1069-1072). IEEE.
- Vohra, S. M., & Teraiya, J. B. (2013). A comparative study of sentiment analysis techniques. *Journal JIKRCE*, 2(2), 313-317.
- Prasad SS, Kumar J, Prabhakar DK, Pal S (2015) Sentiment classification: an approach for Indian language tweets using decision tree. In: International conference on mining intelligence and knowledge exploration. Springer, pp 656–663



23. Venugopalan M, Gupta D (2015) Sentiment classification for Hindi tweets in a constrained environment augmented using tweet specific features. In: International conference on mining intelligence and knowledge exploration. Springer, pp 664–670
24. Kumar A, Kohail S, Ekbal A, Biemann C (2015) Iit-tuda: system for sentiment analysis in Indian languages using lexical acquisition. In: International conference on mining intelligence and knowledge exploration. Springer, pp 684–693
25. Sarkar K, Chakraborty S (2015) A sentiment analysis system for Indian language tweets. In: International conference on mining intelligence and knowledge exploration. Springer, pp 694–702
26. Se S, Vinayakumar R, Kumar MA, Soman K (2015) Amrita-cen@sail2015: Sentiment analysis in Indian languages. In: International conference on mining intelligence and knowledge exploration. Springer, pp 703–710
27. Jha V, Manjunath N, Shenoy PD, Venugopal K, Patnaik LM (2015) Homs: Hindi opinion mining system. In: 2nd International conference on recent trends in information systems. IEEE, pp 366–371
28. Phani S, IEST S, Lahiri S, Biswas A (2016) Sentiment analysis of tweets in three Indian languages. In: Proceedings of the 6th workshop on south and southeast Asian natural language processing, vol 1001. pp 93–102
29. Sharma P, Moh TS (2016) Prediction of Indian election using sentiment analysis on Hindi twitter. In: International conference on big data. IEEE, pp 1966–1971
30. Akhtar MS, Ekbal A, Bhattacharyya P (2016b) Aspect based sentiment analysis in Hindi: resource creation and evaluation. In: Proceedings of the 10th international conference on language resources and evaluation, pp 1–7
31. Das A, Bandyopadhyay S (2010) Phrase-level polarity identification for Bangla. Int J Comput Linguist Appl 1(1–2):169–182
32. Ghosal T, Das SK, Bhattacharjee S (2015) Sentiment analysis on (Bengali horoscope) corpus. In: Annual India conference (INDICON). IEEE, pp 1–6
33. Sharmista A, Ramaswami M (2016) Tree based opinion mining in Tamil for product recommendations using R. Int J Comput Intell Inf 6(2):108–116
34. Nair DS, Jayan JP, Rajeev R, Sherly E (2015) Sentiment analysis of Malayalam film review using machine learning techniques. In: International conference on advances in computing, communications and informatics. IEEE, pp 2381–2384
35. Mukhtar N, Khan MA (2017) Urdu sentiment analysis using supervised machine learning approach. Int J Pattern Recogn Artif Intell 32(02):1–15
36. Hegde Y, Padma S (2017) Sentiment analysis using random forest ensemble for mobile product reviews in Kannada. In: 7th international on advance computing conference. IEEE, pp 777–782
37. Kaur A, Gupta V (2014) N-gram based approach for opinion mining of Punjabi text. In: International workshop on multi-disciplinary trends in artificial intelligence. Springer, pp 81–88
38. Mukku SS, Choudhary N, Mamidi R (2016) Enhanced sentiment classification of Telugu text using ml techniques. In: SAAIP@ 25th international joint conference on artificial intelligence, pp 29–34
39. Mukherjee, S. and Bhattacharyya, P. (2013). Sentiment Analysis: A Literature Survey, pp. 1–51. arXiv preprint arXiv:1304.4520.
40. Rani S, Kumar P (2018) A sentiment analysis system for social media using machine learning techniques: Social enablement. Digit Sch Hum.
41. Kaur J, Saini JR (2015) A Study of Text Classification Natural Language Processing Algorithms for Indian Languages. VNSGU Journal of Science & Technology pp 162-167

AUTHORS PROFILE



Jimmy Singla is working as Associate Professor in School of Computer Science and Engineering, Lovely Professional University Punjab. He has completed Ph.D in CSE from IKG Punjab Technical University, Jalandhar



Gazi Imtiaz Ahmad is Research Scholar in School of Computer Applications, Lovely Professional University Punjab.

