International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-12, Special Issue-1, March-2024
https://doi.org/10.55524/CSISTW.2024.12.1.62
Computing for Sustainable Innovation: Shaping Tomorrow's World (CSISTW-2024) on 16th-17th February, 2024
Organized by B.N. College of Engineering and Technology, Lucknow, India

Machine Learning Approaches for Sentiment Analysis in Hindi Text: A Comprehensive Survey

Brajesh Kumar Khare[1], and Dr. Imran Khan[2]

[1]Research scholar, Department of Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India
[2]Assistant Professor, Department of Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India

Correspondence should be addressed to Brajesh Kumar Khare: brajeshkhare234@gmail.com

Copyright © 2024 Made Brajesh Kumar Khare et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Sentiment analysis from Hindi text is a growing area of research, aiming to understand and categorize the emotions expressed in written content in the Hindi language. Because there is a lot of information on the internet in Indian languages like Hindi, Malyalam, Punjabi, Gujrati, Bengali and others, it is very important to study and find useful and important information from this data. This survey paper offers a summary of the latest progressions and challenges in sentiment analysis specifically tailored for Hindi text. There are four main computational intelligence techniques for getting sentiment from hindi text namely Machine Learning, Deep Learning, Lexicon-based, and Hybrid techniques. In this survey paper we concentrate on Machine learning and Deep learning techniques. This paper discusses about sentiment analysis and their levels, different machine learning models with their features and also the whole process for getting sentiment using machine learning. Furthermore, the paper highlights the challenges associated with sentiment analysis in Hindi, such as the lack of standardized resources, code-mixing, and dialectical variations.

KEYWORDS- Sentiment Analysis (SA), Machine Learning(ML), Deep learning (DL)

I. INTRODUCTION

The internet has made impressive advancements, moving from basic websites and portals to include things like blogs, reviews, opinions, and more[1]. Individuals can share their thoughts and opinions on various products, restaurants, movies, books, events, location and more. The sentiments and viewpoints expressed by individuals can be extremely advantageous for companies seeking to enhance their products and services. Observations indicate that people prefer to provide Assessments and views in their own language. Consequently, there is a growing presence of Hindi text on the internet, including blogs, reviews, and recommendations[2]. India is a country with multiple scripts and languages, and one of the widely used scripts is Devanagari. This script is primarily employed for writing Hindi[3].

II. SENTIMENT ANALYSIS

SA, is a computational approach that utilizes NLP, ML, and linguistic techniques to discern and understand the sentiment conveyed in a given piece of text. SA refers to the process of determining the emotional polarity or feelings that a person has towards a particular entity, such as a product, service, or topic. The polarity can be Positive, Negative or Neutral. SA is an NLP task that necessitates the effective application of text mining methods. It is also occasionally referred to as opinion mining. In the initial stages of the analysis process, subjectivity classification is carried out, which involves recognizing the clause within the sentence. Subsequently, the Assertive text is categorized into either positive or negative. Subjectivity classification is useful for indicating whether the text contains a perspective (subjectivity) or not (objectivity)[4]. For instance, "□□ □□ □□□□ □□□□□ □□□□□ □□" is subjective as it provides a clear perspective on a book. On the other hand, the sentence "□□□□ □□ □□□□□ □□□□□□□□□□□□ □□□" is an objective sentence just provide some general information without expressing a personal opinion. In SA, various techniques are employed to analyze and interpret the subjective information present in the text. To fulfill this goal, a variety of techniques are utilized, including machine learning algorithms, deep learning models, and linguistic rule-based approaches. The goal is to understand the writer's or speaker's feelings, opinions, and attitudes toward a particular subject.

Businesses often use sentiment analysis to gauge customer opinions, assess product reviews, and monitor social media to understand how their brand is perceived. Additionally, SA is applied in various other domains, such as observing and tracking activities across various social media platforms, market research, political analysis, customer service, enabling organizations to make informed decisions based on the sentiments expressed in textual data. Table I represents explanation of sentiment analysis in a tabular format.

A. Steps for Sentiment Analysis

1) ***Establish the Goal****:* Clearly define the goal of your sentiment analysis. Determine the specific questions you want to answer or the insights you seek from the analysis.

2) ***Data Collection****:* Gather the text data that we plan to examine, which can include news articles, customer reviews, social media posts or any other type of text that is relevant to our specific objective.

*3) **Data Pre-processing**:* Clean and ready the text data for analysis. This stage might encompass activities like:

- Removing irrelevant characters, symbols, orspecial characters.
- Tokenization: Dividing a text into separate wordsor tokens.
- Removing stop words: Frequently used words thathave minimal impact on the overall sentiment.
- Stemming or Lemmatization: Simplifying words totheir root form.

*4) **Feature Extraction:*** Convert the pre-processedtext into numerical features that can be employed by ML models. Typical methods commprise TF-IDF, bag-of-words,one hot encoding representation or more advanced word embedding methods like word2vec, FastText.

*5) **Model Selection:*** Select a proper ML or DL model for SA. Common algorithms encompass SVM, NB, Logistic Regresssion, K-Nearest Neighbor, support vector machine Random Forest.

*6) **Training the Model:*** Train the chosen model on the labeled dataset. During this training phase, the model acquires the ability to recognize patterns and associations between features and sentiment labels.

*7) **Evaluation:*** Evaluate the effectiveness of the trained model by using a distinct set of labeled data for performance assessment.

Figure 1 provides an overview of the SA procedure.

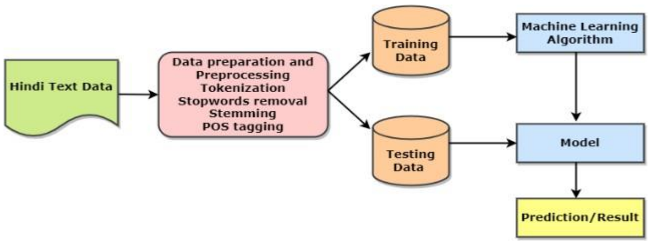

Figure 1: Process of SA

B. ***Sentiment Analysis Approaches*-** Various approaches canbe employed to perform sentiment analysis, and the selection of a specific approach is contingent upon factors such as data characteristics, available resources, and the objectives of the analysis. Some diverse strategies for conducting sentiment analysis includes, ML approach, Deep learning approach, Lexicon approach and Hybrid approach.

C. *Sentiment Levels*- SA can be conducted across three distinct tiers: Aspect level, Sentence level and Documentlevel.

*1) **Aspect-Level SA**:* Analyzing sentiments towards specific aspects or features within a text, such as products, services, or attributes. Example: Assessing customer reviewsto determine sentiments.

*2) **Sentence-Level SA**:* Determining the sentiment expressed in individual sentences within a paragraph or document. Example: Analyzing a product review paragraph and classifying each sentence in different sentiments.

*3) **Document-Level SA***: Evaluating the overall sentiment of an entire document or paragraph. Summarizingthe sentiment of an article or a user review by considering the collective opinion expressed across all sentences in the text. Figure 2 describes the process of SA.

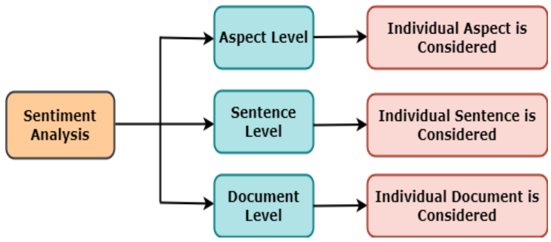

Figure 2: Process of SA

Table 1 describes the various aspects of sentiment analysis.

Table 1 : Explanation Of Sentiment Analysis

Aspect	Description
Definition	SA, is the process of understanding and categorizing sentiments, emotions, or attitudes expressed in Hinditext.
Objective	The primary goal is to identify and characterize sentiments in Hindi text, enabling the analysis of public opinion, customer feedback, or social media content. The focus is on discerning the emotional tone conveyed in the written content.
Text Input	Customer reviews, social media posts, newsarticles, emails, and various textual sources.
Data Pre-processing	Tokenization, Lowercasing, Removing Stop Words, duplicate values, Stemming and Lemmatization, removing punctuations, emojis to prepare text for analysis.
Feature Extraction	TF-IDF, Bag of Words, Word Embeddings, or othermethods to convert text into numerical features for ML models.
Methods	ML Models (Naive Bayes, SVM, KNN etc.) Deep Learning (Neural Network, RNN, LSTM, CNN) Lexicon Based (Rule based), Hybrid approach
Evaluation	Evaluate the model's performance by using metrics like confusion matrices, F1 score, ROC curve, accuracy.
Sentiment Prediction	Apply the trained model to predict sentiment in new,unseen text, assigning labels such as positive, negative, or neutral.

Application	Customer feedback analysis, brand monitoring, social media monitoring, market research, and more.
Challenges	Code mixing, language variations, Cultural Nuances, Resource Scarcity, Linguistic Nuances.
Future Directions	Develop more comprehensive sentiment lexicons and annotated datasets specific to Hindi. Explore domain- specific sentiment analysis tools tailored for industries like healthcare, finance, etc.

III. MACHINE LEARNING MODELS

ML models are like computer tools that are made to help computers learn and figure things out on their own, without someone telling them exactly what to do. They're designed to understand patterns and make predictions or choices without being explicitly programmed for each step. ML models are instrumental in sentiment analysis, especially when dealing with code-mixed languages, where multiple languages are used within the same text.The challenges posed by code-switching and linguistic diversity make machine learning approaches essential for effectively capturing the nuanced sentiment in such contexts. Table II presents a comprehensive examination and description of various machine learning models, offering a broad perspective on the diverse range of models employed in the field.

Table 2: Comprehensive Examination Of Various Machine Learning Mmodels

Model	Use cases	Characteristics	Pros	Cons	Libraries/Tools	Key Hyperparameters
Linear Regression	Regression	Simple, interpretable, linear	Easy to understand	Assumes linearity	Scikit-Learn, Stats models	Regularization parameters (e.g., L1 or L2)
Logistic Regression	Binary Classification	Probabilistic, linear	Simple and efficient	Assumes linear decision boundaries	ikit-Learn, Stats models	Regularization parameters, e.g., C (inverse of regularization strength)
Decision Trees	Classification, Regression	Hierarchical, interpretable	Easy to visualize	Prone to overfitting	Scikit-Learn, XGBoost, LightGBM	Maximum depth, minimumsamples split, criterion
Random Forest	Classification, Regression	Ensemble of decision trees	Improved accuracy, robustness	Computationalexpense	Scikit-Learn, XGBoost, LightGBM	Number of trees, maximumdepth, minimum samples split
Support Vector Machines (SVM)	Classification, Regression	Effective in high-dimensional spaces	Good for complex boundaries	Sensitive to noisy data	Scikit-Learn, LIBSVM	C (regularization parameter), kernel type
k-Nearest Neighbors (k-NN)	Classification, Regression	Instance-based, simple	Simple and easy to implement	Computationally expensive	Scikit-Learn	Number of neighbors (k),distance metric
Naive Bayes	Classification	Assumes feature independence	Simple and computationally efficient	Relies on feature independence	Scikit-Learn	None
K-Means Clustering	Clustering	ioning data into k clusters	Simple and efficient	Sensitive to initial centers	Scikit-Learn	Number of clusters (k), initialization method
Hierarchical Clustering	Clustering	Hierarchy of clusters	No need to specify k	Computationally expensive	Scikit-Learn, SciPy	Linkage method, distancemetric
Neural Networks	Image Recognition, NLP	earning, complex patterns	High predictive accuracy	Requires large data, resources	TensorFlow, PyTorch	Architecture, activation functions, optimizer, learning rate
Convolutional Neural Networks (CNN)	Image Recognition	ized for grid-like data	Effective in image recognition	Computationally expensive	TensorFlow, PyTorch	Architecture, activation functions, optimizer, learning rate
Recurrent Neural Networks (RNN)	Sequential Data	Captures temporal dependencies	Handles sequential data	May suffer from vanishing/exploding gradients	TensorFlow, PyTorch	Architecture, activation functions, optimizer, learning rate
Long Short-Term Memory (LSTM)	Sequential Data	ffective in long-range dependencies	Addresses vanishing gradient problem	Complex architecture, training issues	TensorFlow, PyTorch	Architecture, activation functions, optimizer, learning rate

IV. LITERATURE SURVEY

a) *Machine Learning Techniques-*

Jha et al.[5] proposed the creation of a SA model specifically designed for evaluating reviews of Hindi movies. They employed ML techniques, such as the Naïve bayes Classifier, Support Vector Machine, and ME. The evaluation encompassed a thorough analysis of the classifiers' performance both before and after pre-processing, taking into account both Unigrams and Bigram features. Despite the promising outcomes of their approach, they highlight the importance of conducting further testing on a more extensivedataset.Vandana Jha et al.[6] worked on a Hindi Opinion Mining System for reviews of movies. They used a NB classifier to analyze opinions at the document level. The classification was done by considering the likelihood of a particular category using Unigrams and optimal terms identified using bigram chi-square word features. To make this method better, they suggested reducing errors by using text mining methods and testing different machine learning algorithms to enhance accuracy.

Sharma et al.[7] studied tweets from the 2016 general elections to guess the election results. They used NB and SVM to analyze tweets and figure out people's feelings and support for political parties. Both approaches forecasted that the BJP (Bharatiya Janata Party) would emerge victorious in the elections. However, the study didn't look at emoticons, which are usually important when dealing with tweets.In a study mentioned in [8], four topic areas were already decided: mobile applications, travels, electronics and movies.Detecting these topics was seen as a problem where something could belong to more than one category. They used NB, DT, and SVM as tools to figure this out. The NB classifier worked better for mobile apps and electronics, while DT performed well for movies and travels.Desai et al.[9] worked on finding sarcasm in Hindi sentiment analysis using SVM, a type of tool that sorts sentences into five different groups. To spot sarcastic sentences accurately, you need a good understanding of the meaning and knowledge about the topic.Bafna et al.[10] predicted the type of Hindi poems, placing them into four groups: Updesh-Geet, Desh Bhakti Bhajan and Baal-Geet. They used different methods like NB, DT, Neural Network ,and SVM for this. The classification was done on a collection of 697 poems. SVM yielded superior outcomes in comparison to all alternative methods.Khandelwal et al.[11] worked on predicting the gender of authors in content that was a mix of Hindi and English. They used SVM, Random Forest and NB Classifiers for this. Theyfound some interesting things in the tweets they studied, such as females using more punctuation and hashtags than males.Nanda et al. [12] applied SVM and random forest to analyze Hindi movie reviews. While they achieved good accuracies, they suggested that the results could be even better if neutral cases were considered in addition to positive and negative ones.Yadav et al.[13] used neural networks to analyze a combination of Hindi words. They applied this approach to different areas such as movies, tourism, current affairs, technology, and more. Their study involved sentiment analysis at various levels, comparing the results. They pointed out differences in accuracy when translation was and wasn't considered. Additionally, they proposed that investigating different training algorithms and transfer functions could be apotential focus for future research.Kaushika Pal et al.[14] applied NB and SVM to perform poems classification , utilizing Part-of-speech characteristics and emotional attributes. They emphasized the importance of addressing morphological variations and recommended increasing the dataset size, currently comprising only 55 poems, for more reliable results.Venugopalan et al.[15] investigated SA on Hindi tweets in a restricted setting and put forward a model to tackle the challenges in extracting sentiment from Hindi tweets. The model exhibited moderate effectiveness, achieving a cross- validation accuracy of approximately 56% on the training data and a test accuracy of 43%.Yaman Kumar et al.[16] introduced the BHAAV, which isrecognized as the initial and most extensive Hindi text corpus. It comprises 20,304 sentences extracted from 230 diverse short stories, encompassing 18 genres including Inspirational and Mystery. Three native Hindi speakers assigned each sentence to one of five emotion categories (joy, anger, sadness, suspense and neutral). They applied different classification models, including Logistic Regression, SVM, Random Forests, CNN, and BiLSTM, to the BHAAV dataset. The highest accuracy achieved was 62%, and it was obtainedusing Logistic Regression.Table 3 represents different ML approaches used by researchers.

Table 3: Machine Learning Appraoches

Reference	Model used	Domain	Result
[5]	NB, SVM, Maximum Entropy	Movie Reviews	NB:91.5 SVM:99.5 ME:100
[6]	NB	Movie Reviews	87.3%
[7]	SVM, NB	Ppolitical Parties tweets	SVM 78.4% NB: 62%
[8]	NB, DT, SVM	Electronics, Mobile Applications Travel Movies	Electronics:54.48% Mobile:47.95% Travel:65.25% Movies:91%
[9]	SVM	Movie Reviews	83%
[10]	DT NB, SVM, NN	Hindi poems	SVM: 96% NN:88% NB: 92%
[11]	SVM, NB and Random Forest	Tweets of Demonetization, GST, Triple Talaq	NB:85% SVM:89.5% RF:88.5%
[12]	SVM and random forest	Online ironic andnon - ironic tweets	F1-score SVM: 77% RF: 72%
[13]	Neural Network	Fitness, Enterprise, News and Events, Travel, movie.	80.47%
[14]	NB, SVM	Poems	Non Specific
[15]	SVM DT	SAIL Hindi tweets	43%
[16]	LR SVM Random Forest	Short stories	62%

b) *Deep Learning Techniques-*

Akhtar et al. [17] tackled the challenge of data sparsity by introducing a bilingual word embedding derived from a parallel corpus. They developed a bilingual embedding by training it on a dataset of reviews of products on Amazon. This corpus comprised parallel sentences in both Hindi and English. They developed a Skip Gram word2vec model to represent the semantic closeness between English and Hindi word embeddings. Furthermore, they developed a Bilingual Skip Gram model to generate word embeddings for both Hindi and English languages.Sujata Rani et al. [18] carried out SA on Hindi film critiques employing Convolutional Neural Networks(CNNs) and obtained outstanding results. They proposed extending this approach to explore its effectiveness in other domains. During their experiments, Half of the dataset was used for training CNN models, while the remaining half was utilized for testing. They experimented with different parameter configurations for all the CNN models and noted that the performance of CNNs exceeded that of baseline ML algorithms. Notably, the CNN model with two convolutional layers proved to be the most efficient, attaining an accuracy of 95%.Punit et al.[19] used a CNN approach to identify offensive tweets. They presented a novel dataset named the HEOT dataset and tackled the classification challenge by employing transfer learning, showcasing successful outcomes. The study involved comparing HEOT's performance without transfer learning to that with transfer learning. The authors recommended the application of rrefinement techniques, such as gradient boosting, to improve results.Sane et al.[20] offered a solution for humor detection by combining attention-based bidirectional LSTM and CNN. They employed Word2vec and the fastestSG model to generate bilingual word embeddings. The authors suggested that the results could have been improved by creating aligned multi-lingual word embeddings.Santosh et al. [21] constructed two models, namely Hierarchical LSTM model with attention and the LSTM model for the purpose of Identification of hate speech. They evaluated the outcomes in comparison to SVM and Random class classifiers. Hierarchical LSTM model achieved the best recall and F1 score. The authors proposed that enhancing the capture of additional semantic information, typically essential for detecting hate speech, could lead to improved results.Singhal et al. [22] tackled the issue of multi-lingual sentiment classification using DL. Their method included utilizing publicly available word embeddings and word polarities from English, implementing sentiment classification through CNN. They performed experiments predominantly on reviews of Hindi movies and tourism. The results of their suggested method surpassed those of SVM and other CNN baseline approaches. The system also demonstrated efficacy in dealing with certain unfamiliar words.Siddhartha Mukherjee [23] Employed Long Short-Term Memory (LSTM) for the analysis of mixed Hindi–English code, incorporating both character and word features. The work also outlines a method for building a corpus to train the word embedding. The suggestion is to explore the application of advanced DL models, such as BERT in Hindi–English mixed code language.Table 4 describes the deep learning approach used by different researchers.

Table 4:Deep learning approach

Reference	Model used	Domain	Result
[17]	RNN	Product Reviews	76.29%
[18]	CNN	Movie Reviews	95%
[19]	CNN	HEOT tweet dataset	83.9%
[20]	RNN	Tweets	73.6%
[21]	RNN	Twitter hate and non-hate tweets	66%
[22]	CNN	Movie and Tourism Reviews	Movie: 80.2% Tourism: 88.9%
[23]	LSTM	Movie Reviews	69.84%
[24]	RNN	SemEval 2016 dataset	66.9%

V. CHALLENGES

Analyzing sentiment in code-mixed languages through machine learning encounters numerous challenges, which is shown in figure 4.

*1) **Contexual Problem:*** Ambiguities in language, sarcasm, and implicit sentiment expressions can be challenging for models to decipher accurately.

*2) **Lack of Standarization:*** Hindi text often lacks standardization in terms of spelling and grammar, making it challenging to create a comprehensive sentiment analysis model. Variations in writing styles, transliterations, and informal language use contribute to this issue.

*3) **Code Mixing**:* Hindi speakers frequently engage in code-mixing, where they mix Hindi with English or other languages in their communication. This can complicate sentiment analysis, as the model needs to be able to handle multilingual input and understand the sentiment expressed inboth languages.

*4) **Slang and Informal Language**:* Hindi text often includes slang and informal language, which may not followconventional grammatical rules. Sentiment expressed through such language may be challenging for models to interpret accurately.

*5) **Diverse Vocabulary**:* Hindi has a rich and diverse vocabulary, and people express sentiments using various idioms, metaphors, and regional colloquialisms. Understanding these nuances is crucial for accurate sentimentanalysis, but it adds complexity to the task.

*6) **Limited Labeled Datasets**:* Building a robust sentiment analysis model requires large, labeled datasets for training. However, compared to languages like English, theremay be fewer labeled datasets available for Hindi, making it challenging to train models effectively.

*7) **Limited Resources**:* Compared to English, there may be fewer resources, tools, and pre-trained models available for sentiment analysis in Hindi. This lack of resources can hinder the development and performance of sentiment analysis models for Hindi text.

*8) **Negation Handling**:* Hindi, like many languages, expresses negation in various ways. Identifying negations accurately is crucial for sentiment analysis, as the sentiment of a statement can change drastically with the presence of negation. Models need to be able to recognize negations and their scope within a sentence.

Figure 3 outlines various kinds of challenges that researchers encounter when attempting to analyze sentiments in Hindi text.research suggests that we're moving in a positive direction. Researchers are working on creating better models that consider context, aiming to grasp sentiments more profoundly in this expressive language.

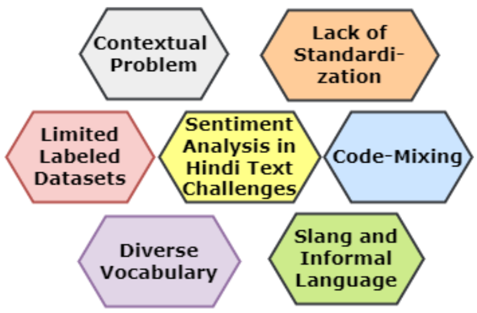

Figure 3: Challenges in SA

VI. CONCLUSION

In conclusion, sentiment analysis in Hindi text presents a compelling area of research with several challenges and opportunities. Researchers, by looking at past studies, have learned a lot about the specific language features of Hindi. They are making progress in figuring out how to deal with these challenges. As the digital landscape continues to evolve, sentiment analysis in Hindi text holds great potential for applications in diverse domains, including social media monitoring, customer feedback analysis, and opinion mining. In summary, even though understanding emotions in Hindi text is tough due to complex language issues, the ongoing.

CONFLICT OF INTERREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends® in information retrieval 2.1– 2 (2008): 1-135.10.1561/1500000001.

[2] Dr. Vivek Kumar Singh. 2015. A survey of opinion mining research. In Hindi Language, International Journal of Advanced Scientific Research & Development 02, 03 Ver. II, (2015), 135– 138

[3] Narang, Sonika Rani, et al. "On the recognition of Devanagari ancient handwritten characters using SIFT and Gabor features." Soft Computing 24 (2020): 17279-17289.

[4] Sharma, Richa, Shweta Nigam, and Rekha Jain. "Opinion mining in Hindi language: a survey." arXiv preprint arXiv:1404.4935 (2014).

[5] Vandana Jha, N. Manjunath, P. Deepa Shenoy, and K. R. Venugopal. 2016. Sentiment analysis in a resource scarce language: Hindi. International Journal of Scientific & Engineering Research 7, 9 (2016), 968–990

[6] Jha, Vandana, et al. "Homs: Hindi opinion mining system." 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS). IEEE, 2015.

[7] Sharma, Parul, and Teng-Sheng Moh. "Prediction of Indian election using sentiment analysis on Hindi Twitter." 2016 IEEE international conference on big data

[8] (big data). IEEE, 2016.

[9] Akhtar, Md Shad, Asif Ekbal, and Pushpak Bhattacharyya. "Aspect based sentiment analysis: category detection and sentiment classification for Hindi." Computational Linguistics and Intelligent Text Processing: 17th International Conference, CICLing 2016, Konya, Turkey, April 3–9, 2016, Revised Selected Papers, Part II 17. Springer International Publishing, 2018.

[10] Prof. Nikita Desai and Anandkumar D. 2016. Sarcasm detection in Hindi sentences using support vector machine. International Journal of Advance Research in Computer Science and Management Studies 4, 7 (2016), 8–15.B

[11] Bafna, Prafulla B., and Jatinderkumar R. Saini. "On exhaustive evaluation of eager machine learning algorithms for classification of Hindi verses." International Journal of Advanced Computer Science and Applications 11.2 (2020).

[12] Khandelwal, Ankush, et al. "Gender prediction in english-hindi code-mixed social media content: Corpus and baseline system." Computación y Sistemas 22.4 (2018): 1241-1247.

[13] Nanda, Charu, Mohit Dua, and Garima Nanda. "Sentiment analysis of movie reviews in hindi language using machine learning." 2018 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2018.

[14] Yadav, Mukesh, and Varunakshi Bhojane. "Semi-supervised mix- Hindi sentiment analysis using neural network." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2019.

[15] Pal, Kaushika, and Biraj V. Patel. "Model for classification of poems in Hindi language based on ras." Smart Systems and IoT: Innovations in Computing: Proceeding of SSIC 2019. Springer Singapore, 2020.

[16] Venugopalan, Manju, and Deepa Gupta. "Sentiment classification for Hindi tweets in a constrained environment augmented using tweet specific features." Mining Intelligence and Knowledge Exploration: Third International Conference, MIKE 2015, Hyderabad, India, December 9-11, 2015, Proceedings 3. Springer International Publishing, 2015.

[17] Kumar, Yaman, et al. "Bhaav-a text corpus for emotion analysis from Hindi stories." arXiv preprint arXiv:1910.04073 (2019).

[18] Akhtar, Md S., et al. "Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality." Association for Computational Linguistics, 2018.

[19] Rani, Sujata, and Parteek Kumar. "Deep learning based sentiment analysis using convolution neural network." Arabian Journal for Science and Engineering 44 (2019): 3305-3314.

[20] Mathur, Puneet, et al. "Detecting offensive tweets in hindi-english code-switched language." Proceedings of the sixth international workshop on natural language processing for social media. 2018.

[21] Sane, Sushmitha Reddy, et al. "Deep learning techniques for humor detection in hindi-english code-mixed tweets." Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2019.

[22] .Santosh, T. Y. S. S., and K. V. S. Aravind. "Hate speech detection in hindi-english code-mixed social media text." Proceedings of the ACM India joint international conference on data science and management of data. 2019.

[23] Singhal, Prerana, and Pushpak Bhattacharyya. "Borrow a little from your rich cousin: Using embeddings and polarities of english words for multilingual sentiment classification." Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016.

[24] Mukherjee, Siddhartha. "Deep learning technique for sentiment analysis of hindi-english code-mixed text using late fusion of character and word features." 2019 IEEE 16th India Council International Conference (INDICON). IEEE, 2019.

[25] Akhtar, Md Shad, et al. "Language-agnostic model for aspect- based sentiment analysis." Proceedings of the 13th International Conference on Computational Semantics-Long Papers. 2019.