International Journal of Speech Technology (2020) 23:757–766
https://doi.org/10.1007/s10772-020-09730-x

Natural Language Processing Based Sentimental Analysis of Hindi (SAH) Script an Optimization Approach

Hewan Shrestha[1] · Chandramohan Dhasarathan[1] · Shanmugam Munisamy[2] · Amudhavel Jayavel[3]

Received: 11 March 2020 / Accepted: 19 June 2020 / Published online: 9 July 2020

Abstract
Sentimental analysis is one of the most common applications of Natural Language Processing (NLP). Sentiment analysis, the term itself refers to identify the emotions and opinions of people through written text. It is concerned with information extraction from any text based on the polarity in social behavior whether it may be positive, negative or neutral. This paper presents a practical dynamic approach on to find the polarity of any sentence and analyse the opinion of the particular sentence. The proposed Sentimental Analysis of Hindi (SAH) script have adopted two different classifier Naïve Bayes Classifier and Decision Tree Classifier is used for the text extraction. The positive, neutral and negative result validation shows a comparative result of sentimental analysis.

Keywords Sentimental analysis of hindi (SAH) · Natural language processing (NLP) · Naïve bayes classifier (NBC) · Decision tree classifier (DTC)

1 Introduction

Natural Language Processing, abbreviated as NLP, is a subfield of linguistics, computer science and artificial intelligence concerned with interactions between computers and human(natural) languages. Natural Language Processing (NLP) is used to apply machine learning algorithms to text and speech.NLP can be used to create systems like sentimental analysis, speech recognition, document summarization, machine translation, spam detection, named entity recognition, question answering, auto complete, predictive typing, information extraction, and so on. Nowadays, NLP is used to power search engines, filter spam and to obtain analytics in a fast and scalable manner. Sentimental analysis is the first NLP task that every Data Scientist needs to perform to understand the working mechanism and necessity of data in NLP. The article is organized with various sections it discusses an overview of sentimental analysis and its deviations by different authors is illustrated under Sect. 2. The proposed approach is enhanced in Sect. 3 with various pseudo normality and tokenization. In Sect. 4 Practical Approach for the Sentiment Analysis evaluation is illustrated. At Sect. 5 consists of discussion about the necessity of the proposed approach and its advantages. Finally the paper is concluded with sufficient and addressable limitations in sentimental analysis.

2 Literature Study

María Lucia et al. (2020), The Internet allows everyone to share information between peoples in form of text, visual or audio and allows users to give their views and opinions regarding different products, brands, and more. Those

✉ Chandramohan Dhasarathan
pdchandramohan@gmail.com

Hewan Shrestha
shresthahewan12@gmail.com

Shanmugam Munisamy
shaninfo247@gmail.com

Amudhavel Jayavel
info.amudhavel@gmail.com

[1] Department of Computer Science & Engineering, Madanapalle Institute of Technology & Science, Madanapalle, India

[2] Department of Computer Science and Engineering, Vignan's Foundation for Science Technology & Research (Deemed to be University), Guntur, Andhra Pradesh, India

[3] School of Computing Science & Engineering, VIT Bhopal University, Bhopal, India

opinions that users express can have great influence on suppliers of products, educational firms and so on.

Mahmoud Al-Ayyoub et al. (2018), Sentiment Analysis (SA) is a field of research that lies in between deep learning, natural language processing and machine learning. It is mainly concerned with the automatic extraction of opinions provided in a particular text. Many researchers have beenoccurred in the field of Sentiment Analysis mainly on English texts, while other languages such as Arabic is having less attention.

Vandana Jha et al. (2017), Sentiment Analysis being a sub area of Natural Language Processing (NLP) refers to extracting user's views and classifies it with respect to the polarity of the text provided. To annotate the corpora in every possible domain of interest, we created a sentiment aware dictionary using multiple domain data using labelled data from source domain and unlabeled data from both source and target domains. The work is actually carried out in Hindi which is the official language of India.

Asma Mekki et al. (2018), In recent years, Hindi language has become the primary focus of research in terms of Natural Language Processing. Nanada kishore et al. (2011) proposed a web service based suitability to assessment for cloud deploying it for computing in a betterment of communication system. Dhasarathan et al. (2018) have designed a transmission line interfacing to synchrophasor signals. Moreover, network speech anomaly detection by deep learning approach as a comprehensive novel.

Cabot et al. (2019), In the past decades in the biomedical field as well as other fields including social media, entity recognition has been studied extensively. With the time, many advanced algorithms have been used to improve the entity recognition in formal medical text for English language. However, adapting these algorithms to free text is quite difficult as they are designed for formal texts like biomedical texts.

Zakariae et al. (2019), The actual user search content has been neglected by the rapid growth of electronic data and wide expansion of the World Wide Web (WWW). Information retrieval and its interactions have been received more attention over the past decades due to that constraints imposed by that fact.

Atoum et al. (2020), This paper proposes a QinU framework (QinUF) in order to measure QinU competently by receiving software reviews. The framework is composed of three different components which includes QinU prediction, polarity classification, and QinU scoring.

Mohammad et al. (2019), In the field of Natural Language Processing (NLP), Sentiment Analysis is considered to be one of the most attractive research area nowadays. The main focus of sentiment analysis is to recognize the opinions and emotions of the users over the written texts. However, it is still considered a challenging topic which are yet not solved such as modern accents, slang words, spelling and grammatical mistakes.

Song et al. (2020), Having wide number of applications, Natural Language Processing is widely used for Sentiment Analysis, which consists of various areas including statistics, linguistics, psychology, and artificial intelligence. To identify the views of users, NLP gets and evaluative factor which may be positive or negative towards a subject topic, person, or idea. There are two main categories of automatic sentiment analysis which includes lexicon-based approaches and machine-learning approaches.

Wei et al. (2019), For the past few decades, Sentiment Analysis has been a popular field of research in Natural Language Processing. Most of the current researches on sentiment analysis focus on identifying explicit sentiments. For that reason, analyzing implicit sentiments has been one of the most difficult tasks in sentiment analysis. In this article, a BiLSTM model with multipolarity orthogonal attention has been proposed for implicit sentiment analysis.

Gan et al. (2020), Sentiment Analysis is one of the important fundamental task of Natural Language Processing (NLP) which aims to distinguish sentiment polarity i.e. positive, negative or neutral, towards a particular target entity. One of the major challenge in sentiment analysis is to model the interaction between the specified target entity and its context.

Li et al. (2020), Sentiment analysis is one of the branch of linguistic computing research, which is used to distinguish texts as positive, negative or neutral. Some existing methods of affective sentiment analysis can be divided into three categories which includes knowledge-based techniques, statistical methods, and hybrid methods. Some common sources of sentiment words includes Affective Lexicon, linguistic annotation scheme, WordNet-Affect, SentiWordNet, SenticNet.

Zhuang et al. (2019), In the era of digitalization, a large number of people share their views and opinions on the World Wide Web (WWW) such as online reviews. There are a number of other people with around 73–87% who use those online reviews for the choice of purchasing the product. Many researches on sentiment analysis are mainly focused on machine-learning based. However, it is also having a problem, as they need a huge amount of training data to perform well.

Mowlaei et al. (2020). Sentiment Analysis approach is mainly focused towards detecting the views and opinions of written reviews so that to have a better understanding of public opinion. One of the common approaches in sentiment analysis is to use the lexicons to generate features for classification of reviews. Manimaran et al (2020), proposed an deep learning approach for speech technology to identify the true negative and false positive ratio in identifying speech process. Ashok kumar et al. (2018) propose an fault test analysis to identify the transmission with the help of synchrophasor signals. Dhasarathan et al. (2018), proposed an natural inspired approached for healthcare management system. Alami Merrouni et al. (2019), proposed an approach for contextual information retrieval and its trends to apply in currentscenario.

Chandramohan et al. (2015), illustrated the process of collecting and storing data in a confidential process, it would protect the user information privacy for a pervasive & ubiquitousenvironment a multi-agent approach.

Chandramohan et al. (2018, 2019), A coordinator-specific privacy-preserving model for e-health monitoring using artificial bee colony approach is proposed to ensure the users data anddesigned a Fog enabled secure and privacy obfuscation for IoT services.

Estrada et al. (2020), proposed an opinion mining and emotion recognition applied to learning environments which leads to academia and research field for an effective teaching learningprocess.

Hassonah et al. (2019), illustrated an efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter. Knowledge-Based Systems.

Jha et al. (2018), proposed a novel sentiment aware dictionary for multi-domain sentiment classification approach for the betterment of natural language processing.

Ruz et al. (2020), an approach for effective sentiment analysis of Twitter data during critical events through Bayesian networks classifiers which would support the social media dataanalysis.

3 Proposed Sentimental Analysis of Hindi (SAH) Script

Sentiment Analysis aims to estimate the sentiment of text based on its context. The sentiment of any text can be defined as a value that says whether the expected opinion is positive or negative. Using the techniques from Natural Language Processing (NLP), sentiment analysis field looks at users' expressions and in turn associate emotions with what the user has provided. Sentiment Analysis has become key component to systematically extract, identify, and quantify the data.

NLP based analysis is more ruminate in sentimental analysis. The future research progress towards artificial intelligent and machine learning approaches. It is essential to adopt NLP approach to solve the problem in a systematic process and it is applied with necessary recommendations. The proposed approach would be applicable for gifted child activity analysis and giving support to encourage them to have a normal life among the society.

Sentiment Analysis for Hindi script is hardly found which has immense possibilities it tends to revolutionize the surveys and review collections in Hindi with its growing applicability to a wide variety of applications from computer service to marketing. Hindi Sentiment Analysis generally consists of sample Hindi data collection, data processing, feature extraction, and classification. Feature extraction aims to detect and extract features that can be used to determine the meaning of a given Hindi-Contextual-Data. The extracted features should be able to classify the data reliably into positive, negative or neutral class.

4 Augmentation of Sentiment Analysis in English and Hindi languages

It is well-known that different languages have their own unique ways of expression. The basic difference between English and Hindi language is the language structure. For example, English language has S-V–O (Subject-Verb-Object) structure, while Hindi language follows S–O-V (Subject-Object-Verb) structure. The basic structural difference between English and Hindi language has consequences in deciding the polarity of text. The same set of words with slight variations and changes in the word order affect the polarity of the words in the text. Therefore, a deeper linguistic analysis is required while dealing with Hindi language to perform Sentiment Analysis. For example, consider the sentences given below which demonstrates the difference between language structure of English and Hindi language.

English: Shyam is playing football.
S V O
Hindi: श्याम फुटबॉल खेलरहा है।
S O V

Trained Dataset validation of positive and negative SVO is observed in Table 1. It is tested with all scenarios for its effectiveness. The observation of SAH procedures are marked and get validated in all aspects for the benefits of SAH. The pesudo code is tested with SAH under all aspects to conform the necessity and applicability under various field of research.

5 SAH Procedures with Pesudo Codes

Sentiment Analysis of any text can be performed by the following five steps.

Step 1 Tokenization.
Step 2 Cleaning the data (Removing punctuation).
Step 3 Removing the Stop Words.
Step 4 Classification.
Step 5 Calculation.

6 Expressability of SAH

Step 1: Tokenization.

Tokenization is the process of defining a paragraph into different set of statements or dividing a statement into different set of words.

Table 1 Trained dataset date to validate the positive and negative SVO

1	Ted	Politicians do not have permission to do what needs to be done	राजनीतज्ञोंकेपासजोकार्यकरनाचाहएि, वहकरनेकअिनुमतनिहींहै
2	ted	I'd like to tell you about one such child	मईआपकोऐसेहीएकबच्चेकेबारेमेंबतानाचाहूंगी
3	indic2012	This percentage is even greater than the percentage in India	यहप्रतशितभारतमेंहनि्दुओंप्रतशितसेअधकिहै।
4	ted	what we really mean is that they're bad at not paying attention	हमयेनहींकहनाचाहतेकवि्रोध्याननहींदेपाते
5	indic2012	The ending portion of these Vedas is called Upanishad	इन्हींवेदोंकाअंतमिभागउपनषिदकहलाताहै।
6	tides	The then Governor of Kashmir resisted transfer, but was finally reduced to subjection with the aid of British	कश्मीरकेतत्कालीनगवर्नरनेइसहस्तांतरणकावरिोधकयिाथा,लेकनिअंग्रेजोंकीसहायतासेउनकीआवाजदबादीगयी
7	indic2012	In this lies the circumstances of people before you	इसमेंतुमसेपूर्वगुज़रेहुएलोगोंकेहालातहैं।
8	ted	And who are we to say, even, that they are wrong	औरहमहोतेकौनहैंयहकहनेभीवालेकविगलतहैं
9	indic2012	""Global Warming "" refer to warming caused in recent decades and probability of its continual presence and its indirect effect on human being	ग्लोबलवॉर्मगि्सेआशयहालहीकेदशकोंमेंहुईवार्मगिऔरइसकेनरिंतरबनेरहनेकेअनुमानऔरइसकेअप्रत्यक्षरूपसेमानवपरपड़नेवालेप्रभावसेहै।
10	tides	You may want your child to go to a school that is not run by the LEA—a non-maintained special school or an independent school that can meet your child 's needs	होसकताहैकआिपचाहतेहोंकआिपकानऋृनमेनटेनडहयबनिाकसिीसमर्थनकेहुवशिषस्कूल,याकसिीस्वतंत्रस्कूलमेंजाए, इजसकेपासवशिषशैक्षणकिजऋृऋृरतोंवालेबच्चोंकेप्रतसिहूलयितहो
11	tides	Please ensure that you use the appropriate form	कृपयायहसुनश्चितिकरलेंकआिपसहीफॉर्मकाप्रयोगकररहेहैं
12	indic2012	Category: religious text	श्रेणी:धर्मग्रन्थ
13	indic2012	This period summarily is pepped up with devotion	यहकालसमग्रतःभक्तभिावनासेओतप्रोतकालहै।
14	ted	So there is some sort of justice	तोवहाँन्यायहै
15	tides	The first two were found unreliable and the prosecution case rested mainly on the evidence of the remaining five approvers	पहलेदोकोअवश्विसनीयमानकरबाकीपांचमुखबरिोंकेआधारपरमुकदमाचलायागया
16	tides	They had justified their educational policy of concentrating on the education of a small number of upper and middle-class people with the argument that the new education would gradually ' filter down …	कमसंख्यावालेउच्चएवंमध्यमश्रेणीकेलोगोंतकहीअपनीशकि्षानीतकिोकेंद्रतिकरनेकोइसतर्ककेसाथन्यायसंगतबतायाकनियीशकि्षाZक्रमश:ऊपरसेनीचेकीओरछनतेहुएजायेगी
17	indic2012	And now at present the naturecure, Ayurvedic and modern treatments are taking place through the government in Nepal	हालमेंनेपालकेहस्पतालसामन्यतयाआयुर्वेद, प्राकृतकिचकित्सिातथाआधुनकिचकित्साकरकेसरकारीसेवावद्यिमानहे।
18	indic2012	Parliament time frame is 5 years and this will be dissolved before that	लोकसभाकीकार्यावधि 5 वर्षहैपरंतुइसेसमयसेपूर्वभंगकयिाजासकताहै
19	tides	ii Register Courts, empowered to try causes for amounts not exceeding Rs 200, when authorised by the judges	रजस्टिरन्यायालयजन्हिेंन्यायाधीशद्वाराप्राधकिृतकएिजानेपर 200 रु.तककेवादोंकानर्णियकरनेकाअधकिारथा
20	indic2012	Extreme weather due to increased mortality; displacements and economic loss will be compounded through growing population. Although, temperate climate has some benefits out of it such as decreased mor…	बढ़तीहुईमौतों displacements औरआर्थकिनुकसानजोकीअतविाद्रीमौसम (extreme weather) केकारणसंभावतिहैबढतीहुईजनसंख्या (growing population)केकारणऔरभीबदतरहोसकतेहैं. हालांकशिीतोष्णक्षेत्रमे…
21	tides	Of these Lahadi is a popular one	' लाडड़ीलोकप्रयिस्त्रीनृत्यहै
22	tides	Even a concentration of 0.001 ppm of hydrogen sulphide in the water can emit the smell of rotten egg	यहांतककपिानीमेंहाइड्रोजनसल्फाइडकी 0.001 पीपीएममात्रासेभीसड़ेहुएअंडेकीबदबूआतीहै
23	indic2012	Islam is the world's second-largest religion, after Christianity	इस्लामधर्म (الإسلام) ईसाईधर्मकेबादअनुयाइयोंकेआधारपरदुनयिाकादूसरासबसेबड़ाधर्महै।
24	ted	This changed slowly	धीरेधीरेयेसबबदला
25	tides	Far more interesting are genetic diseases that arise essentially from the mutation of a single gene allowing a simple Mendelian distribution to appear in the offspring concerned	एकहीजीनकेउत्परविर्तनकेकारणहोनेवालेआनुवंशकिरोगइनसभीरोगोंसेअधकिमहत्वपूर्णहैं.मेंडेलकेसरलवतिरणकेनयिमकेअनुसारयहजीनसंतानोंमेंप्रकटहोताहै
26	tides	The FIs are expected to offload the stake in favour of Suzuki -LRB- currently 50 per cent stakeholder -RRB- and to the general public later	बाद्मेंवतिैइयसंस्थाएंइनशेयरोंकोसुजुकी(फलिहाल 50 प्रतशितशेयरधारक) औरआमजनताकोबेचदेंगी
27	ted	Were being produced	उत्पन्ननहींकजातीथी

Table 1 (continued)

1	Ted	Politicians do not have permission to do what needs to be done	राजनीतिज्ञोंकेपासजोकार्यकरनाचाहिए, वहकरनेकेअनुमतनहींहै
28	tides	Naren had three or four meetings with the Consul but found that he was making no progress	नरेंद्रनेकौंसुलिसेतीनयाचारमुलाकातेकीकंतुउन्होनेंमहसूसकयिाकविकसिीभीप्रकारसेबातआगेनहीबढ़ापारहेहै
29	tides	All those who stand today under this flag are Indians, not Hindus, not Muslims, but Indians	जोभीइसझंडेकेनीचेखड़ाहैवहनहंदूहै,नमुसलमान, बल्कविहहंदुस्तानीहै
30	tides	In some cases the High Courts have the jurisdiction to deal with matters more effectively as courts of first instance than the Supreme Court	कुछमामलोंमेंप्रथमस्तरकेन्यायालयहोनेकेकारणउच्चन्यायालयोंकोउच्चतमन्यायालयकीअपेक्षामामलोंपरअधिकिकारगरढंगसेकार्रवाईकरनेकीअधिकारिताहै
31	ted	And you can see, this LED is going to glow	औरजैसाआपदेखरहेहै, येएल.ई.डी. जलउठेगी।
32	tides	Under the Act more than one Claims Tribunal can be set up and their territorial jurisdiction is specified by notification constituting the Tribunals	अधनियिमकेअधीनएकसेअधकिदावाअधकिरणस्थापतिकएिजासकतेहैंऔरउनकीअधकिारिताकोअधकिरणोंकेगठनकीअधसिूचनामेंवनिरि्दष्टकरदयिाजाताहै
33	ted	to turn on the lights or to bring him a glass of water	लाईटजलानेकेलएियाउनकेलएिपानीलानेकेलएि,
34	indic2012	Maine	मेन
35	indic2012	IN farsi philosophy firdaus garden is presented as ideal one in Mugal literatures	फारसीरहस्यवादमेंमुगलकालीनइस्लामीपाठ्यमेंफरिदौसकोएकआदर्शपूर्णताकाबागबतायागयाहै।
36	ted	Can you imagine saying that?	क्याआपयेकल्पनाकरसकतेहै
37	tides	Head lice need to maintain contact with a host in order to survive	जंदिारहनेकेलएिजूँकसिीजीवकेसम्पर्कमेंरहनाआवश्यकहै
38	ted	Three: this is a good road in—right near where our factory is located	तीसरी: येहमारीफै़क्ट्रीकेपासकीएकअपेक्षाकृतबेहतरसडकहै।
39	tides	This may include injunctions, which can be highly effective preventive measures which can reduce the nuisance caused whilst allowing people to keep their homes	इसमेंकानूनीरोक(इंजक्शन) शामलिहैजोबहुतहीप्रभावकारीरोककाउपायहैजोपैदाहुएउपद्रवकोकमकरसकताहैऔरइसबीचलोगोंकोअपनेघरोंमेंरहनेकीअनुमतभीहोतीहै
40	ted	What's going on?"	क्याहोरहाहैये?"
41	indic2012	But after his demise, his bother Sardar Vallabh Bhai Patel did not accept this will and filed a case in the court	मगरउनकेनधिनकेपश्चातउनकेभाईसरदारखल्लभभाईपटेलनेइसवसीयतकोस्वीकारनहींकयिाऔरउसपरअदालतमेंमुकदमाचलाया।
42	tides	He also saw with great anguish how, following the ultra-left line of the sixth Comintern congress, the communists had destroyed themselves and gravely damaged the left movement	उन्होनेंबड़ेदुखकेसाथयहभीदेखाककैसेछठीकमनिट्रनकांग्रेसकीअत्विामपंथीमार्गअपनाकरसाम्येवादयिोंनेंअपनेकोनष्टकरलयिाथाऔरवामपंथीआंदोलनकोभीबढ़तेदेखाथा
43	ted	There are also financial reforms in rural China	ग्रामीणचीनमेंआर्थकिनवीनीकरणहुयेहैं।
44	tides	That is the final lesson of the story	यहइसकहानीकाअंतमिसबकहै
45	tides	It is a clash all along, of the old with the new, of real politik with idealism, of the means with-the end, of love claimed as of right and love given of free will, of home-bred virtue with the w…	इसमेंटकरवाहीटकरावहै—नएकेसाथपुरानेका, राजयसत्ताबनामआदर्शवादका, साधयऔरसाधनका, अधकिारपूरणप्रेमऔरप्रेमकीस्वतंत्रताका, औरबाहरीजंगलीहवाकेसाथग्राहस्थकियाघरेलूमूल्योंक…
46	tides	Progressive-minded, he advocated widow remarriage and raising the marriageable age of girls	अपनेप्रगतशिीलवचिारोंकेकारणउन्होनेंवधिवा-ववािहऔरकन्याओंकीववािहयोग्यआयु-सीमाबढ़ानेकीभीवकालतकी
47	tides	The range was comprehensive, the analysis sober and lucid, and the style so charming that what he said was itself afine specimen of literature	वशि्लेषणऔरवविचनबड़ेगंभीरऔरसरसथेऔरशैलीइतनीमोहकथीकसिारा-का-सारालेखनसाहत्यिकाउत्कृष्टनदिर्शनबनगया
48	tides	Twenty of the years from 1860 to 1908 were years of famine	सन् 1860 और 1908 केबीचके 20 वर्षअकालकेवर्षरहे
49	indic2012	This position is similar to armchair. In this the male puts its legs slightly up near the knees. and his hands are helpful to support the hips of the female. The female spreads her leg touching his ba…	-झूलायहहत्थाकुरसीसेमलितीजुलतीपोजीशनहैइसमेंपुरुषअपनेपैरोंकोघुटनोंकेपाससेथोड़ाउपरउठालेताहैऔरउसकेहाथमहलिाकोकूल्होंकेपाससेसहारादेनेकाकामकरतेहैंतथामहलिाअपनेदोनोंपै…

```
Python 3.7.6 (default, Jan 19 2020, 22:34:52)
[GCC 9.2.1 20200117] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> from nltk.tokenize import word_tokenize
>>> sentence = 'तुम्हा रा  ना म क्या  हे?'
>>> result = word_tokenize(sentence)
>>> print(result)
['तुम्हा रा ', 'ना म', 'क्या ', 'हे', '?']
>>>
```

Step 2: Cleaning the data (Removing punctuation).

```
Python 3.7.6 (default, Jan 19 2020, 22:34:52)
[GCC 9.2.1 20200117] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import string
>>> text = dict((ord(char), None) for char in string.punctuation)
>>> s = 'मुझे क्ल @ मेरी  किता ब $ मि ली ।'
>>> result = s.translate(text)
>>> print(result)
मुझे क्ल  मेरी  किता ब  मि ली ।
>>>
```

Cleaning the data means to remove those special characters/words which do not add any value to the analytics part of the sentence.

Step 3: Removing Stop Words.

```
Python 3.7.6 (default, Jan 19 2020, 22:34:52)
[GCC 9.2.1 20200117] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> from nltk.tokenize import word_tokenize
>>> from nltk.corpus import indian
>>> s = 'कृपया  जा कर अपने मि त्र को फोन करें।'
>>> stop_words = set(indian.words('hindi_stopwords.txt'))
>>> word_tokens = word_tokenize(s)
>>> sentence = [w for w in word_tokens if not w in stop_words]
>>> sentence = []
>>> for w in word_tokens:
...     if w not in stop_words:
...             sentence.append(w)
...
>>> print(word_tokens)
['कृपया ', 'जा कर', 'अपने', 'मि त्र', 'को', 'फोन', 'करें।']
>>> print(sentence)
['कृपया ', 'जा कर', 'मि त्र', 'फोन', 'करें।']
>>>
```

Removing those words which do not add much value to the analytics part of the sentence.

Step 4: Classification.

We can model our data with bag of words or lexicons, which are dictionary of pre-classified set of words.

Step 5: Calculation.

This is the final step in the sentiment analysis of any text. This step corresponds to finding the final sentiment score of the sentence. In general, if polarity > 0, then the sentence is positive and the sentence is negative if the polarity < 0. However, we can get the polarities in fractional values depending on the procedure applied.

7 SAH-1Pseudo-code

```
function wordlist(words)
    return dict ([(word, True) for word in words])

function main()
    select the  nltk.classify.util
    retried nltk.NaiveBayesClassifier from nltk.classify
    incorporate  nltk.DecisionTreeClassifier from nltk.classify
    import names from nltk.corpus

    positive_words ← "positive.txt"
    negative_words ← "negative.txt"
    train_set ←positive_words + negative_words
    classifier ← train(train_set)
```

8 SAH-2 Pseudo-Code

```
pos ← 0
neg ← 0
sentence ← "some input"
for id in sentence do
        result ←classifier.classify(data)
        if result = 'positive'
                pos ← pos + 1
        if result = 'negative'
                neg ←neg +1
display pos and neg
```

9 Practical Dynamic Approach for the Sentiment Analysis

The dynamic approach for the evaluation of Sentiment Analysis of Hindi text is given by various pseudocode and it is tested with various functionality.

SAH-3 **Pseudo-code:**

```
function wordlist(words):
        return dict ([(word, True) for word in words])
function main:
        select nltk.classify()
        select word_tokenize from nltk.tokenize
        incorporate nltk.NaiveBayesClassifier from nltk.classify
        retrieve nltk.DecisionTreeClassifier from nltk.classify
        select names from nltk.corpus
        import indian from nltk.corpus
                positive_vocab← open("positive.txt").read()
                negative_vocab← open("negative.txt").read()
        pos_wordlist = [(wordlist(pos), 'pos') for pos in positive_vocab]
        neg_wordlist = [(wordlist(neg), 'neg') for neg in negative_vocab]
        train_set←pos_wordlist + neg_wordlist
        classifier ←nltk.classify.NaiveBayesClassifier.train(train_set)
        neg ← 0,  pos ← 0, entry ← "some input";
        sentence ←entry.lower(), stop_words← set(indian.words("hindi_stopwords.txt"))
        word_tokens←word_tokenize(sentence), filtered_sentence← [w for w in word_tokens if
not w instop_words],  filtered_sentence← []
        for w in word_tokens do:
                if w not in stop_words do:
                        filtered_sentence← append(w)
        for id in filtered_sentence do
                result ←classifier.classify(wordlist(id))
                if result = 'pos' do:
                        pos ← pos + 1
                if result = 'neg' do:
                        neg ←neg+ 1
        print(pos)
        print(neg)
```

The result can be seen in the tables given below for both positive and negative sentences separately as well the combined version of both positive and negative sentences. The article incorporated with a set of new quality parameters for the identification of sentimental words and phrases for the betterment of analysis. False positive and true negative parameter is added in addition to check the minute hot spot of the analysis and it is updated in the article.

In Table 2, Fig. 1 it is observed that there is a variance in positive and negative value by both the classifiers. It also supports in all means with derived algorithms. The dynamic approach values are tabulated for better result observation.

Validation of positive sentence is illustrated in Table 3 and the same is shown in Fig. 2. It is observed that the negative sentence identification by the classification approach is tested under various inputs. Moreover, it is shows a progressive improvement in all aspects.

Amalgamation of both positive and negative sentences has all valid contribution by testing in various aspects and it is illustrated in Table.4 and the same is shown in Fig. 3. In sentimental analysis more research have been carried out in pin pointing the reorganization of words in different forms. It is not sufficient for the analysis. So, in this article it is focused on NLP based sentimental analysis for the identification true and false positive words. It would help the identification process more significant and appropriate.

Table 2 Tested with negative sentences

No. of SAH X_i in e^{x_i}	Naïve Bayes classifier		Decision tree classifier	
	+ ve	− ve	+ ve	− ve
$X_{1\text{-}4}$	0.0	1.0	0.0	1.0
$X_{2\text{-}8}$	0.0	1.0	0.3333	0.6666
$X_{3\text{-}12}$	0.0	1.0	0.25	0.75
$X_{4\text{-}16}$	0.1	0.9	0.3	0.7
$X_{5\text{-}20}$	0.0833	0.9166	0.3333	0.6666

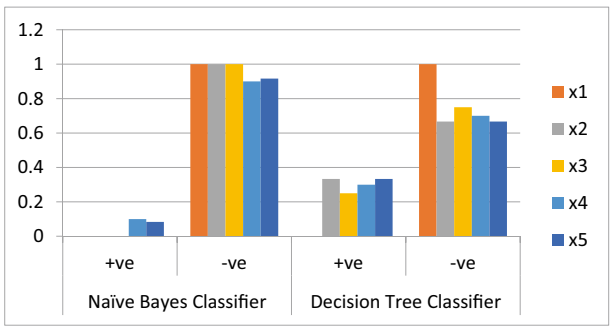

Fig. 1 Illustration of negative sentences

Table 3 Validation and testing of positive sentences

No. of SAH X_i in e^{x_i}	Naïve Bayes classifier		Decision tree classifier	
	+ ve	− ve	+ ve	− ve
$X_{1\text{-}4}$	0.5	0.5	0.5	0.5
$X_{2\text{-}8}$	0.5	0.5	0.6666	0.3333
$X_{3\text{-}12}$	0.375	0.625	0.75	0.25
$X_{4\text{-}16}$	0.4	0.6	0.7	0.3
$X_{5\text{-}20}$	0.30769	0.69230	0.52846	0.461538

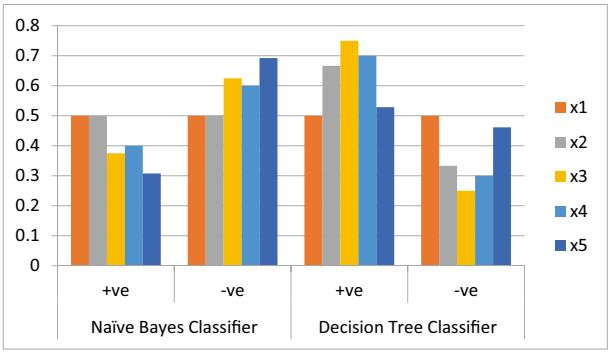

Fig. 2 Positive sentences

Table 4 Testing the combination of positive and negative sentences

No. of words e^{x_i} where X_i	Naïve Bayes classifier		Decision tree classifier	
	+ ve	− ve	+ ve	− ve
X_1-4	0.3333	0.6666	0.6666	0.3333
X_2-8	0.14285	0.85714	0.57142	0.428571
X_3-12	0.10	.9	0.5	0.5
X_4-16	0.0769	0.92307	0.53846	0.461538
X_5-20	0.1875	0.8125	0.5625	0.4375

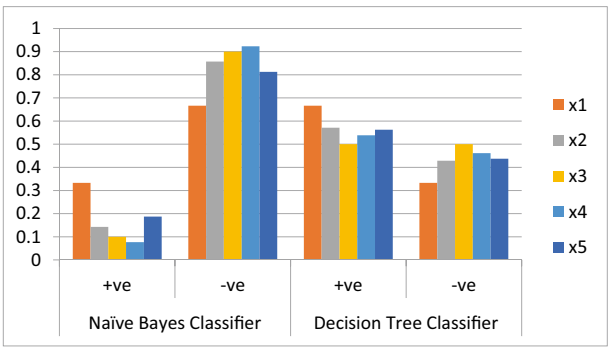

Fig. 3 Amalgamation of positive and negative sentences

10 Discussion

For the sentiment analysis of Hindi sentence and reviews, this paper used a dataset of hindi stop words. The dataset is created with collective of positive and negative wordlists as illustrated in Table 1. To test the data collected is tested and proceeded using the NLTK (Natural Language Toolkit) library which is readily available for Python programming language. By importing all required functions and features from NLTK library, the proposed SAH successful validated to get the required analysis which is illustrated in the result section. Moreover, the trained datasets is classified by using NaiveBayesClassifier and DecisionTreeClassifier which is verified for the analysis of positive and negative words to segregate the appropriate mapping.

11 Conclusion

In this paper, a dynamic approach is proposed to find sentiment analysis of Hindi text using Natural Language Processing. Nowadays, social media and news feed are common with Hindi text as well and it would become difficult in some situations for the peoples who are unfamiliar to the Hindi language. Most researchers are focused towards information extraction from those texts available over the internet. Many authors have given their approach towards analysing sentiment from any text. However, in this article it is tested by providing a practical approach for analyzing the sentiment of Hindi text for those researchers who are willing to research on Indo-Aryan culture and scripts.

References

Alami Merrouni, Z., Frikh, B., & Ouhbi, B. (2019). Toward contextual information retrieval: A review and trends. *Procedia Computer Science, 148*, 191–200.

Ashok Kumar, S., & Chandramohan, D. (2018). Fault test analysis in transmission lines throughout interfering synchrophasor signals. *Elsevier- ICT Express*. https://doi.org/10.1016/j.icte.2018.03.003.

Atoum, I. (2020). A novel framework for measuring software quality-in-use based on semantic similarity and sentiment analysis of software reviews. *Journal of King Saud University: Computer and Information Sciences, 32*(1), 113–125.

Cabot, C., Darmoni, S., & Soualmia, L. F. (2019). Cimind: A phonetic-based tool for multilingual named entity recognition in biomedical texts. *Journal of Biomedical Informatics, 94*, 103176.

Chandramohan, D., Manimaran, A., Reddy, R., & Tripathi, D. (2019). Fog enabled secure and privacy obfuscation for IoT services. *Journal of Advance Research in Dynamical & Control Systems., 11*(8), 1604–1610.

Chandramohan, D., Rajaguru, D., Vengattaram, T., & Dhavachelvan, P. (2018). *A coordinator-specific privacy-preserving model for e-health monitoring using artificial bee colony approach*. Hoboken: Wiley. https://doi.org/10.1002/spy2.32.

Chandramohan, D., Sathian, D., Rajaguru, D., Vengattaraman, T., & Dhavachelvan, P. (2015). A multi-agent approach: To preserve user information privacy for a pervasive & ubiquitous environment. *Egyptian Informatics Journal (Elsevier), 16*, 151–166. https://doi.org/10.1016/j.eij.2015.02.002.

Chandramohan, D., Vengattaraman, T., & Dhavachelvan, P. (2016). A new privacy preserving technique for cloud service user endorsement using multi-agents. *Elsevier-Journal of King Saud University-Computer and Information Sciences, 28*(1), 37–54. https://doi.org/10.1016/j.jksuci.2014.06.018.

Estrada, M. C. B., Cabada, R. Z., Bustillos, R. O., & Graff, M. (2020). Opinion mining and emotion recognition applied to learning environments. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2020.113265.

Gan, C., Wang, L., Zhang, Z., & Wang, Z. (2020). Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis. *Knowledge-Based Systems, 188*, 104827. https://doi.org/10.1016/j.knosys.2019.06.035.

Hassonah, M. A., Al-Sayyed, R., Rodan, A., et al. (2019). An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter. *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2019.105353.

Jha, V., Savitha, R., Deepa Shenoy, P., Venugopal, K. R., & Sangaiah, A. K. (2018). A novel sentiment aware dictionary for multi-domain sentiment classification. *Computers & Electrical Engineering, 69*, 585–597.

Kishore, N. M. S., Jayakumar, S. K. V. (2011). Web service suitability assessment for cloud computing. In Wyld, D. C. (Ed.), Trends in network and communications LNCS NeCoM/WeST/WiMoN 2011, CCIS 197, 2011 (pp. 622–632). Berlin:Springer.

Li, W., Qi, F., Tang, M., & Yu, Z. (2020). Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2020.01.006.

Manimaran, A., Chandramohan, D., Shrinivas, S. G., & Arulkumar, N. (2020). A comprehensive novel model for network speech anomaly detection system using deep learning approach. *International Journal of Speech Technology*. https://doi.org/10.1007/s10772-020-09693-z.

Mekki, A., Zribi, I., Ellouze, M., & Hadrich Belguith, L. (2018). Critical description of TA linguistic resources. *Procedia Computer Science, 142*, 230–237.

Mowlaei, M. E., Abadeh, M. S., & Keshavarz, H. (2020). Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2020.113234.

Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2020.01.005.

Song, C., Wang, X.-K., Cheng, P., Wang, J., & Li, L. (2020). SACPC: A framework based on probabilistic linguistic terms for short text sentiment analysis. *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2020.105572.

Wei, J., Liao, J., Yang, Z., Wang, S., & Zhao, Q. (2019). BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. *Neurocomputing, 383*, 165–173. https://doi.org/10.1016/j.neucom.2019.11.054.

Zhuang, L., Schouten, K., & Frasincar, F. (2019). SOBA: Semi-automated ontology builder for aspect-based sentiment analysis. *Journal of Web Semantics, 60*, 100544. https://doi.org/10.1016/j.websem.2019.100544.