An Empirical Analysis on Various Techniques Used to Detect the Polarity of Customer Satisfaction in Sentiment Analysis

A.Ilavendhan[a], Soumya Ranjan Jena[b] and S.N. Manoharan[c]

[a]Assistant Professor, Department of CSE, School of Computing, Vel Tech RangarajanDr.Sagunthala R&D Institute of Science & Technology, Chennai, Tamil Nadu, India(ORCID: 0000-0001-9241-120X)
[b] Assistant Professor, Department of CSE, School of Computing, Vel Tech RangarajanDr.Sagunthala R&D Institute of Science & Technology, Chennai, Tamil Nadu, India (ORCID: 0000-0003-1099-5649)
[c]Associate Professor, Department of CSE, School of Computing, Vel Tech RangarajanDr.Sagunthala R&D Institute of Science & Technology, Chennai, Tamil Nadu, India (ORCID: 0000-0003-3144-490X)
[a]ilavendhans@gmail.com, [b]soumyajena1989@gmail.com, [a]kritimanoharan@gmail.com

_____

Abstract: Sentiment analysis is an emerging application of NLP (Natural Language Processing). This is also called opinion mining or attitude detection. In the text mining field, Sentiment Analysis is continuous area of research. It is a procedural treatment of attitude, feelings and textual content. The fundamental thought is to discover text polarity and order it as positive, negative, or neutral. It supports human to take a good judgment. This survey paper gives an extensive summary of the previous updates in this area. Several currently proposed algorithms and numerous upgrades to different SA applications have been investigated and summed up in this review. These articles are classified by their commitment to different SA techniques. Areas identified with SA (business monitoring, polarity observation, social media monitoring) that has recently attracted researchers is discussed. The fundamental objective of this survey is to provide a complete picture of SA practices and associated fields with brief explanation. The significant role of this study includes a refined classification of current papers and a depiction of ongoing patterns in sentiment analysis and research in its associated fields.

Keywords: Learning, Sentiment analysis, Feature selection, Sentiment classification, Emotional detection

_____

1.   Introduction

Sentiment analysis is also known as Opinion mining. It tends to be highly valuable in trading field, social media fields, and so forth. At the point when an organization sells an item, it is significant to understand how client's response to the item for additional development or development of their business. Not simply as far as business, we have frequently observed that person give feedback in the form of reviews of films.The main goal is to know whether the user has a negative, positive and neutral view of a product or anything else. Number of paper related to sentiment analysis increases at high rate. There is a lot of growth in analytics these days. It is one of the fastest growing research fields. It is observed that the public opinion, their motivation is of a political nature. The period of mid 2000s was when Sentiment examination has magnified. The companies were mainly focused Product reviews which are available on the web. There is now much research work on emotion analysis to identify human attitudes, reactions, or thoughts.

The Internet has become a major part of our lives today. Before purchasing any product from e-commerce sites, the customer first checks the validity of a positive, negative or neutral product. They likewise utilize these websites to realize what other client's assessment about the item are. This is why mining and sentiment extraction of these data has become an important area of research.

There are two terms polarity and subjectivity which can be viewed as parts of emotional examination. Sentiment examination covers the extent of work at the sentence level, document level and sub-sentence level.

i)   Subjectivity/Objectivity-To investigate sentiment we should initially distinguish text which is either subjective or objective. Iftext has emotions means it is subjective text otherwise text is objective because it contains only factual details.

a)   Subjective: Tanhaji is a super film. (This line has a feeling (super); along these lines it is subjective).

b)   Objective: Om Raut is a maker of Tanhaji (This line has no feelings, it is a reality, hence it is classified as objective).

_____

ii) Polarity- Text which is subjective means it can be categorized into three classifications dependent on the sentiment in the text.

a) Positive: I like iPhone 11 mobile because picture feature of this telephone is better.

b) Negative: Battery backup is poor.

c) Neutral: I usually remain hungry until early afternoon (this sentence is subjective because it contains user thoughts and feelings but is neutral because it has no positive or negative polarity.)

iii) Sentiment Analysis- It can also be carried out at various level:

a) Sentence Level Analysis: At this stage, the task is to decide whether each sentence expresses feelings. This stage indicates target sentences that express real data and subjective opinions. In this case, the treatments are two-fold; First find out whether the sentence indicates an opinion or not, and then evaluate the opinion polarity. The fundamental difficulty originates from the fact that target sentences lead to sentiments.

b) Document Level Analysis: At this stage the main aim is to prescribe the entire view of the document. At the document stage, SA display that every document indicates fillings on an individual entity.

c) Aspect Level Analysis: At this stage, we can do better investigation and this stage requires the usage of natural language processing. At this stage, the description of feelings can be done through polarity and the objective of opinion. In this matter, the remedies are two-fold: in question we first recognize the elements and aspects of the unit, and afterwards on each aspect access the response.

2. Methodology

To carry out sentiment analysis, we must follow a few stages: data collection, data preprocessing, Feature selection and sentiment ^classification. As the classification, a general plan of sentiment analysis is appeared in Figure 1. Every single step of work methodology has its significance.

Data Collection

We need data first for work with sentiment analysis, so we first collect customer reviews on a given product from e-commerce sites. The reviews of the customer are stored in a table and then filter the table based on customer review. Check each review stored in table is subjective or objective. If review is objective then discard review from table because it does not contain customer emotional information, otherwise I can be considered in a given format.

Data Pre-processing

We first clean up the data set. This data cleaning procedure is called data preprocessing. At this stage, it is compulsory to eliminate unwanted words used by commas, full stop, colon, or any particular characters or symbols in customer reviews. This type of words do not relate to any sentimental rules. Therefore, it is necessary to eliminate such conditions to reduce further complications. Preprocessing stage in this article is done in two steps as described under:

- In a customer survey, sometimesURL is also described but we well know that URL represent address and it does not contain any emotional information or sentiment value so we delete it.

- Few particular characters or symbols are deleted. Coma, full-stop, colon and semi colon are eliminated.
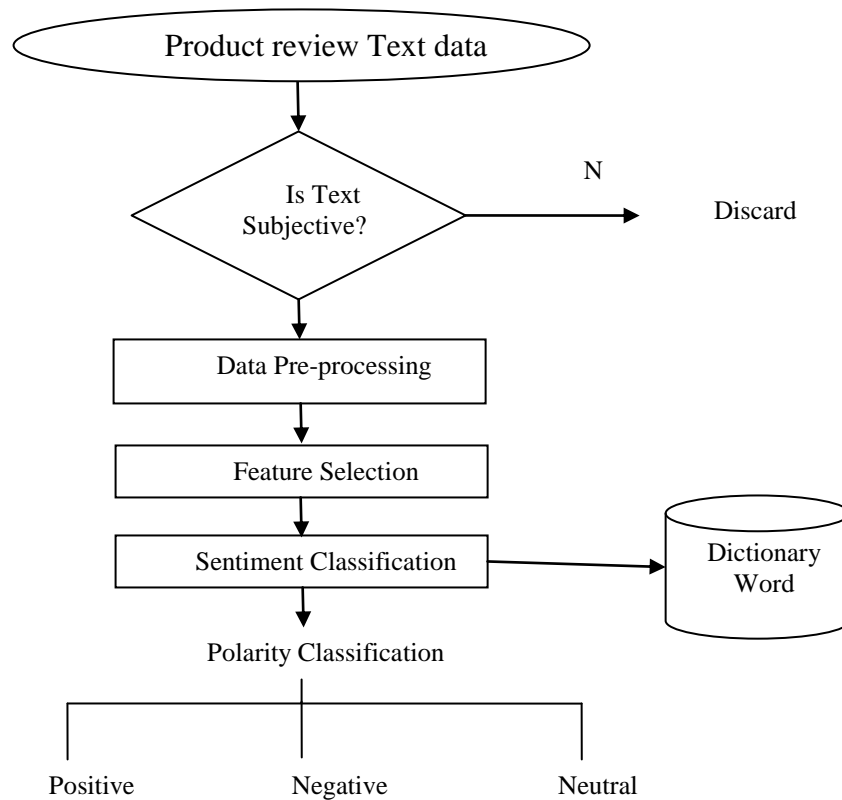
Figure 1. Sentiment Analysis Process

Feature Selection

The first step of the sentiment classification problem is to capture and pick features in text. Few of the recent highlights are given as follows:

Terms presence and term frequency: It refers to single words or word n-gram and their frequency counts. It returns the weighted binary word (means whether the word present or not if present then it returns to zero otherwise it returns to one) or used the frequency weights term which refers to the relative weightage of the features.

POS taggin*g:* Searching for adjectives, these are significant sign of feelings.

Opinion words and phrases*:* These are generally words which indicate feelings such as like or hate, bad or good. On the other side, a few expressions indicate feelings without the utilization of opinion words.

Negations Handling: The existence of negative words does not equate with good or bad but can change opinion. For instance: the film was not bad has a word 'bad' in which it is negative but 'not' invert the polarity to positive.

Stemming: Stemming is assists to reduce locality in words. For instance, words such as 'jumping', and 'jumped' can be stemmed to 'jump'.

Stop words-These are the English words which does not give knowledge about sentiments. There are several stop words which lists are available on the internet. This is useful to pick them during the pre-processing stage. For instance: Prepositions (in, near, beside), Pronouns (he/she, it), Articles (a, the) are ^stop words.

Sentiment Classification Technique

Sentiment classification methods comprehensively arranged into ML based methods, lexicon based methods, and hybrid based methods. The ML based methods implement the popular ML algorithms and utilizes language features. TheLB approach is depends on the sentiment lexicon, a group of familiar and preempted emotional words. It is classified in corpus based and dictionary based approaches which use semantic or statistical methods respectively for predicting the emotions polarity. Two approaches such as (ML based + LB based) combine to

make a hybrid approach, which plays a significant role in most of the methods. The most popular algorithms of different approaches and SCs are described in Figure 2 as previously stated.

With the help of machine learning techniques, Text classification can be broadly subdivided in to supervised or unsupervised learning methods.Supervised method applies on labeled training document. When these labeled training documents are hard to evaluate then we use unsupervised methods.

To examine the text, we use lexicon based methods which is depending on searching the opinion dictionary. There are two ways to deal with this procedure such as dictionary based and corpus based methods. In a dictionary-based method, this is based on searching for seed words and then discovers their synonyms and antonyms. But in corpus based methods, it starts with aseed list of opinion words, and then discovers different words in the bigger corpus to assist them detect the words with a particular trend. The next subsections will include a brief overview of the algorithms and related articles of the two approaches. Figure 2 illustrating the various methods used in sentiment analysis.
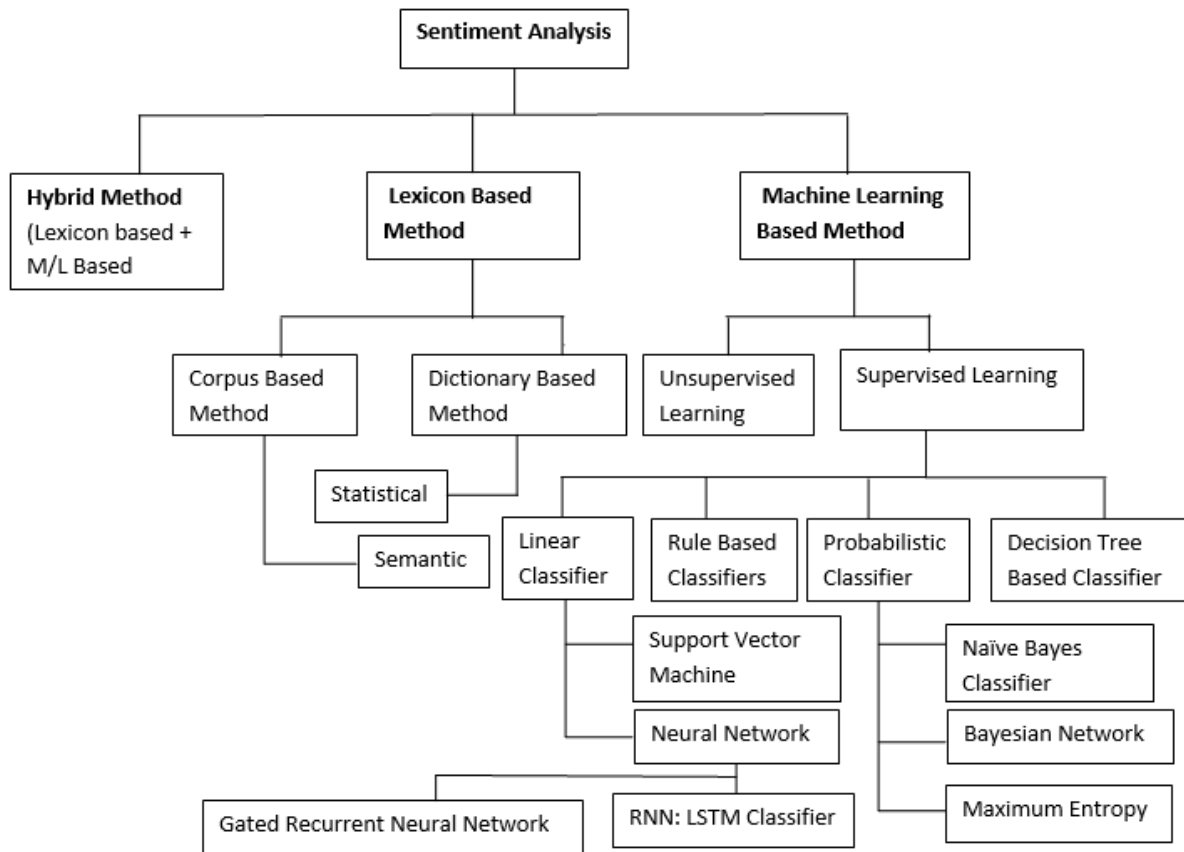

Figure 2. Sentiment Analysis Classification Technique

Machine Learning Approach:

It is known as classification technique. Classification is executed using text properties. Properties draw out from the text. Two types of ML based approach are: Supervised and unsupervised learning.

Supervised Learning

These strategies based on the presence of labeled training documents. Several type of supervised classifications methods shown below in a table 1. In the following subsections, we discuss few of the most common classifier which we can use in sentiment analysis.

Probabilistic Classifiers:

In machine learning, these classifiers utilize mixed models for classification. The composite model expectsthat all class is a portion of the mixture. Every composite component is a simple design, which produces the opportunity to sample a particular word for that part. Such classification is also known as generative classification. The three most popular classifications are discussed in subsequent subsections.

Naive Bayes Classifier:

NB classification is the easiest and highly usually used classification. It calculates the posterior probabilities of a class which is depend on the words distribution in the document. It relies on Bayes' theorem which determine the probability of a given class (Q) with a given feature set, represented by vector R= $(r_1, r_2 \dots \dots, r_n)$. Bayes Theorem gives a way to find out the conditional probability:

$$P(Q \mid R) \quad \frac{P(Q)*P(R \mid Q)}{P(R)}$$

Where

- P (Q) =Prior probability of a specific class.
- P (R |Q) is the predictor probability when class given.
- P(R) = Predictor prior probability.
- P (Q| R) = Probability of class when predictor given.

This Classifier supposes that everyfeatureis independent and the above equation can be represented as:

$$P(Q \mid R) = P(Q)* \frac{\sum_{i=1}^{n} P(R_i \mid Q)}{P(R)}$$

Linear Classifier

It classifies properties which is depends on linear combination value**.** Let V= {v1, v2, v3….} is linear coefficient vector and U= {u1, u2, u3….} is word frequency, and the outcome of linear predictor is given below when S be a linear scalar.

LIP=U**.**V+S.

This predictor is known hyper plane which divide two classes.

Support Vector Machine**:** SVM can be utilized for classification. Its primary aim is to separate the better linear partition for classification. This is a classification method that is not probabilistic. For a given training set, labeled for each class, the SVM training algorithm creates a model that assigns new data to one or two classes. Hyper plane is utilized to isolate the two classes. In figure 3 under, for instance, to predict diamond and rectangle shapes we estimate three hyper planes P, Q and R. R is best partition because on both sides items are at highest distance from hyper plane and Q is very bad partition.
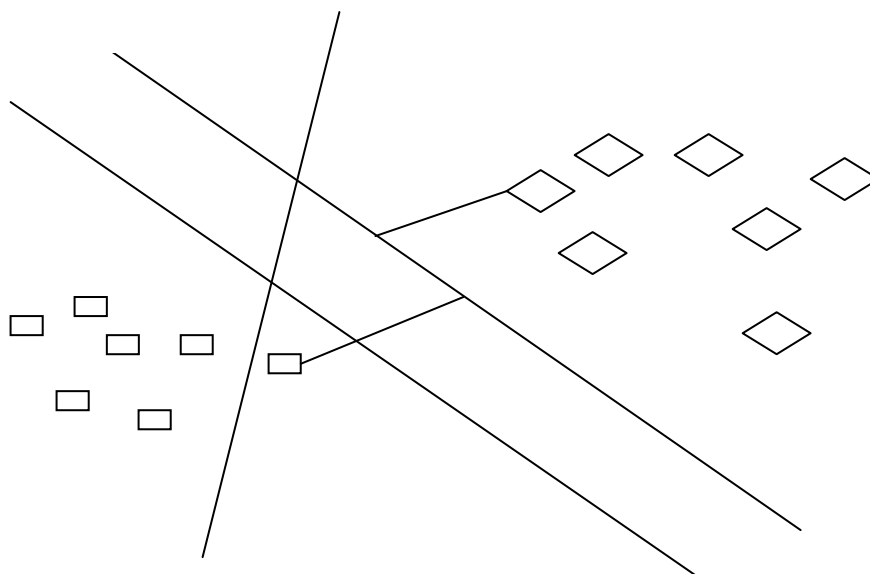

Figure 3. Classification Problem using SVM

Decision Tree Classifier

It is another type of non-parametric supervised method. This classifier is utilized for both classification and regression task. The aim is to generate a model which evaluates the values of the target variable by learning simple decision rules derived from data attributes. In this classifier, a condition is utilized to divide the data. The data will be placed in one class if satisfy condition otherwise rest of the data put in the other class. It is a repetitive process. There are several partitions: a single attribute partition that searches for the ^presence or absence of a specific word to classify, Similarity depends multi-attribute partition that matches the predefined words of a given document word to perform classification and Discriminate depend multi-attribute partition use discriminates to perform partition.

Rule based Classifier

In this classifier, data space is made up of a set of rules. The feature set conditions appear on the left side, while class label appears on right side. The conditions are on the term existence. The term absence is utilized rarely because it does not include information on small data. This classification uses few rules as IF, THEN. It can be written as:

IF condition1 THEN decision1

During the training phase the rule can be generated depends on requirement.

Table 1 show the differentiation between the thirteen papers and recognize the usage of every technique in sentiment analysis.

Table 1. Use of Sentiment Analysis Techniques

Year	Approach	Techniques used	Experiment	Dataset	Accuracy	Future Scope
2016	Opinion mining and sentiment analysis	NLP Technique	To obtain the tweets polarity on behalf of sentiment score.	Tweets from twitter	80.6%	The credibility of the tweets being analyzed can be verified by reviewing them from other websites and sources and comparing the two.
2017	Reduced feature based sentiment analysis.	Hybrid approach (lexicon based + machine learning based)	Movie reviews tend to be positive or negative in nature.	Movie Reviews	95.43	In the future, traits depend on intensive learning techniques may be chosen to further enhance emotion based classification.
2017	Customer product review using machine learning.	Naive Bayes, SVM, Decision Tree	Polarity of reviews.	400000 review of 4500 mobile phone taken from Amazon.	The highest value of accuracy obtained for SVM in all three models is 81.75%	To improve performance, we can apply deep learning algorithm like LSTM, GRNN etc. Also predict rating of a product from the reviews
2017	Text based sentiment analysis.	POS Technique, CFG	To check emotion of sentence or paragraph is positive,	To design S/W that takes review of customer in form of	---	In this paper, some CFGs are implemented, but several CFGs can be carried

			negative or neutral.	sentence or paragraph and check polarity.		out to build the software more sensible and powerful.
2018	Rating of reviewers based on text.	Hybrid approach (Lexicon based + M/C Learning (SVM classifier))	Predicting the ratings of films based on the text of the reviews.	Reviews of eight films from Amazon	82%	With the popularity of deep learning, We can test similar procedures using LSTM or CNN and improve accuracy.
2018	Product review based on sentiment analysis and fusion of EEG response.	Random forest based regression techniques, ABC Optimization Algorithm.	To predict customer satisfaction based on product review.	42 Product & Brand review data set, 40 Participant (25 Male, 15 Female) EEG Dataset while watching 42 products.	---	Extract robust feature of EEG Signal by exploring other wavelet transform algorithm to improve the prediction of product.
2018	Microblog Sentiment Analysis	Laplacian regularized least squares	To identify sentiment polarity.	Sentiment Analysis of Twitter data set: HCR and OMD	Structure homogeneity context gets better accuracy among all social contexts. Accuracy of SSC is 79.16%, 78.99 on HCR, OMD dataset respectively	To apply support vector machines (SVMs) for better performance.
2018	ML based sentiment analysis.	Naive Bayes classifier	To predict tweets sentiments are either in positive, negative or neutral.	Tweets from twitter	When sentiments of tweets extracted from WSD sentiment analyzer then we get the highest acuuracy79 %	We can find patterns of political parties based on Twitter.
2018	Nobel adaptable approach for sentiment analysis.	Hybrid Method (lexicon based + M/c learning)	To predict polarity and identify tweets is either positive class or negative class.	Twitter dataset (2016 US presidential election)	90.21%	Proposed approach undergo from few drawbacks. It does not differentiate the level of effectiveness of different matrices for expressing an emotion.

2019	Sentiment analysis of tourist review	LSTM classification algorithm	To predict polarity of tourist this is either positive or negative.	Tourist review	84%, 90% and 98% accuracy with SVM, K-Nearest Neighbor, and LSTM respectively.	The proposed framework can be integrated into multiple classifications with deep learning.
2020	Sentiment analysis using machine learning Algorithm	Sequential minimal optimization with decision tree algorithm	To predict polarity of tweets.	Tweets of airline dataset	89.47%	An efficient optimization technique can be used to improve performance.
2020	Sentiment Analysis Using Gated RNN	Gated Recurrent Neural Network	To predict product review positive, neutral or negative.	Amazon product reviews	70.42, 71.19 accuracy with Bi-LSTM, Bi-GRU respectively so Bidirectional gated recurrent unit (Bi-GRU) perform better than Bi-LSTM.	We can extend the work and gain such firmly rooted knowledge of human interactions such as humor or irony.
2020	Polarity shift in sentiment classification	SVM Classifier	Polarity Classification	Amazon (user reviews)	97%	Future work will focus on addressing more challenging areas such as spam and counterfeiting, negativity and sarcasm.

The mostly technique used in theoretical type is POS tagging and lexicon-based methods but practically there are a lot of ML technique (such asSVM, Naive Bayes) and some deep learning technique like LSTM, GRNN used to classify SA discussed in table 1.

3.   Conclusion & Future Scope

In this paper, this study gives a review on ongoing updates in SA algorithm and applications. Thirteen articles distributed and referred in Table 1 are arranged and summed up. These articles contribute to a number of emotion analysis-related fields that use different emotion analysis techniques in real-world applications. After examination of these papers, it turns out to be evident that the rise of emotion classification and trait selection algorithms is as yet an open region for research.SVM and Naive Bayes are now popular ML techniques which can be employed for the sentiment classification problem. In future, some deep learning algorithm (like LSTM classifier, GRNN) can be utilized to ease the SA process. It assists with delivering increasingly precise outcome in sentiment analysis. Also Future work is an extension of the comparative cycle with continuous new research.

References

Harshali P. Patil, Dr. Mohammad Atique, "Sentiment Analysis for Social Media: A Survey",978-1-4673-8611-1/15/$31.00 ©2015 IEEE.

NurfadhlinaMohdSharef, Harnani Mat Zin, SamanehNadali, "Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data", Journal of Computer Sciences 2016, 12 (3): pp 153.168

Rushlene Kaur Bakshi, Navneet Kaur, Ravneet Kaur, Gurpreet Kaur, "Opinion Mining and Sentiment Analysis", 2016 International Conference on Computing for Sustainable Global Development (INDIACom), 2016, pp 452–455.

DoaaMohey El-Din Mohamed Hussein, "A survey on sentiment analysis challenges", Journal of King Saud University – Engineering Sciences (2018) 30, PNo. 330–338

Sruthi S, Reshma Sheik, Ansamma John, "Reduced Feature Based Sentiment Analysis on Movie Reviews Using Key Terms", 978-1-5386-3864-4/17/$31.00 © 2017 IEEE.

Harpreet Kaur, VeenuMangat, Nidhi, "A Survey of Sentiment Analysis techniques", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017), pp921–925.

ZeeniaSingla, Sukhchandan Randhawa, Sushma Jain, "Sentiment Analysis of Customer Product Reviews Using Machine Learning", 2017 International Conference on Intelligent Computing and Control (I2C2).

Biswarup Nandi, MousumiGhanti, Souvik Paul, "TEXT BASED SENTIMENT ANALYSIS", International Conference on Inventive Computing and Informatics (ICICI 2017), IEEE Xplore Compliant -Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9, pp9–13.

MohammadamirKavousi, SepehrSaadatmand, "Estimating the Rating of Reviewers Based on the Text", First International Conference on Data Analytics & Learning 2018, arXiv:1805.08415v1 [cs.CL].

SudhanshuKumar ,MahendraYadava , ParthaPratim Roy, "Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction", Information Fusion 52 (2019), pp 41–52

XiaomeiZou, Jing Yang, Jianpei Zhang, "Microblog sentiment analysis using social and topic context", PLOS ONE | https://doi.org/10.1371/journal.pone.0191163 February 2, 2018, pp 1-24

Ali Hasan, Sana Moin, Ahmad Karim andShahaboddinShamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts", Math. Comput. Appl. 2018, 23(1), 11; https://doi.org/10.3390/mca23010011

Imane El Alaoui , Youssef Gahi, RochdiMessoussi, YounessChaabi, Alexis Todoskoff and AbdessamadKobi, "A novel adaptable approach for sentiment analysis on big social data" Journal of Big Data (2018) 5:12 https://doi.org/10.1186/s40537-018-0120-0, pp 1-18

Nikhil Kumar Singh, Deepak Singh Tomar, Arun Kumar Sangaiah, "Sentiment analysis: a review and comparative analysis over social media" Journal of Ambient Intelligence and Humanized Computing (2018), https://doi.org/10.1007/s12652-018-0862-8.

Yong Chen, Bin Zhou, Weina Zhang, Wenjie Gong, Guangfu Sun, "Sentiment Analysis Based on Deep Learning and its Application in Screening for Perinatal Depression", 2018 IEEE Third International Conference on Data Science in Cyberspace, pp 451-456.

Priyanka Thakur, Dr. Rajiv Shrivastava, "Sentiment Analysis of Tourist Review using Supervised Long Short Term Memory Deep Learning Approach", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 7, Issue 2, February 2019, ISSN(Online): 2320-9801 ISSN (Print) : 2320-9798, pp 592-604.

PriyankaNamdev, Prof.Lakhan Singh, "A Survey of Sentiment Analysis Process and Technologies", International Journal of Engineering Trends and Technology (IJETT) - Volume 67 Issue 11- Nov 2019, ISSN: 2231-5381, pp 153-156.

PetrBerka, "Sentiment analysis using rule-based and case-based reasoning", Journal of Intelligent Information Systems (2020), https://doi.org/10.1007/s10844-019-00591-8.

A. Naresh, P. Venkata Krishna, "An efficient approach for sentiment analysis using machine learning algorithm", Evolutionary Intelligence (2020), https://doi.org/10.1007/s12065-020-00429-1

Sharat Sachin, AbhaTripathi, Navya Mahajan, Shivani Aggarwal, PreetiNagrath, "Sentiment Analysis Using Gated Recurrent Neural Networks", SN Computer Science (2020) 1:74, https://doi.org/10.1007/s42979-020-0076-y.

Long Mai, Bac Le, "Joint sentence and aspect-level sentiment analysis of product comments", Annals of Operations Research https://doi.org/10.1007/s10479-020-03534-7.

NehaNandal, RohitTanwar, JyotiPruthi, "Machine learning based aspect level sentiment analysis for Amazon products", Spat. Inf. Res. @ Korean Spatial Information Society 2020 https://doi.org/10.1007/s41324-020-00320-2.

SalimaBehdenna, FatihaBarigou, and GhalemBelalem, "Sentiment Analysis at Document Level"© Springer Nature Singapore Pte Ltd. 2016 A. Unal et al. (Eds.): SmartCom 2016, CCIS 628, pp. 159–168, 2016.

Aggarwal Charu C, Zhai Cheng Xiang. Mining Text Data. Springer New York Dordrecht Heidelberg London: _ Springer Science+Business Media, LLC'12; 2012.

Svetlana Kiritchenko ,Xiaodan Zhu ,Saif M. Mohammad Sentiment Analysis of Short Informal Texts. Journal of Artificial Intelligence Research 50 (2014) 723-762. Page no-723.

Arora, Piyush. "Sentiment Analysis for Hindi Language." Diss. International Institute of Information Technology Hyderabad, 2013.

WalaaMedhat, Ahmed Hassan, HodaKorashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal. Page no-1094.

Yelena Mejova, Padmini Srinivasan. Exploring feature definition and selection for sentiment classifiers. In: Proceedings of the fifth international AAAI conference on weblogs and social media; 2011.

NehaNandal, RohitTanwar, JyotiPruthi, "Machine learning based aspect level sentiment analysis for Amazon products", Spat. Inf. Res. @ Korean Spatial Information Society 2020 https://doi.org/10.1007/s41324-020-00320-2.