Sentimental Analysis Using Supervised Learning Algorithms

Devansh Shah
School of Computer Science and Engineering
Lovely Professional University
Jalandhar Punjab, India
devanshshah476@gmail.com

Arun Singh,
School of Computer Science and Engineering,
Lovely Professional University,
Jalandhar,Punjab, India
arunmandiarun2001@gmail.com

Sudha Shanker Prasad
Assistant Professor, School of Computer Science and Engineering
Lovely Professional University
Jalandhar Punjab, India
sudha.21826@lpu.com

2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM) | 978-1-6654-5319-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCAKM54721.2022.9990320

***Abstract* — The humongous growth of online- grounded operations, similar to forums and blogs, led to commentary and reviews related to diurnal conditioning. Sentimental analysis is the process of collection of data and analyzing one's studies, ideas, also extremities about colorful motifs, products, motifs, and services. One's ideas can be useful to companies, governmental organizations, and individualities by gathering data and practicing vision- grounded opinions. Still, emotional analysis and the process of assessment are at stake numerous challenges. These challenges produce walls to directly interpret feelings and determine applicable emotional diversity. Emotional judges and excerpt practical information from the textbook using natural language processing as well as digging the textbook. This composition discusses a complete idea of how to negotiate this work and the use of Sentimental analysis. Also, it is responsible for evaluating, comparing, and investigating the styles used to understand both good and bad. Eventually, Sentimental analysis challenges are explored to define unborn directions.**

Keywords—Opinion Mining, NLP, Supervised Machine Learning, Sentiment Analysis

I. INTRODUCTION

Sentiment or emotion Analysis is one of the trending topics of Natural Language Processing. It deals with the digging of ideas to get the feeling about the colorful motifs being bandied. The introductory function is to divide a piece of textbook into paragraphs that express an idea in a story about one of the two opposing feelings; former perceptivity was espoused to dissect long- term feelings written content similar as letters, emails, reviews etc. People nowadays are very frequent on social media whether it's a post or a comment, uploading stories, making content or uploading pictures is something which makes social media trending and as a result generates humongous data.

Opinion Mining or Sentimental analysis is one of the arising motifs for nearly every assiduity related to products and services. Client reviews, feedbacks about a particular product or a service lays an excellent foundation for a healthy dataset. Social media is one similar source currently, to gain reviews, get feedback, conduct any check, collect any data. Social media platforms, where people tend to note and give their review, it becomes a veritably good source for IT companies to gain data and also dissect about the performance of product, what new variations can be done, in terms of service it can be the major area of enhancement, quality of service, perfecting technology used in the quality and numerous other factors. It's largely trending currently, and is a no way ending content as it's a veritably important tool to test new products and services in the request among common man and also laboriously contribute towards the conservation of the same according to the response generated by druggies to stand in the competitive request. .

II. Literature Review

Many real- world operations bear emotion analysis to further probe, for illustration, product analysis, to determine which corridor or features of a product attract an audience in terms of product design and user interface. Research methodologies of Subhashini etal.(2021)[1] presents a comprehensive review of ultramodern sentiment literature. It tells us how to prize textbook rudiments from sound or uncertain ideas, to represent information in ideas, and to classify it. Mowlaei etal.(2020)[2] raise a dynamic wordbook approach grounded on the emotion-sensitive aspect. (Prasad etal. 2015)[3] describe about indian language tweets. The authors describe two strategies for erecting two dynamic workbooks to help classify feelings according to their characteristics: fine- grounded strategy and inheritable mechanisms.. To analyze each & every review, select a subset of several workbooks in many workbooks. Feelings Analysis has preliminarily been applied in a variety of fields from hospices to flight reservation system and health industry to the stock request(Zvarevashe and Olugbara 2018)[4]. Sentimental analysis worked on hostel reviews to more understand client preferences and dislikes. In comparison(Valencia et al. 2019)[5], it can determine stock request trends and cryptocurrencies grounded on request sentiment. The authors(Ahmad etal. 2019)[6] dissect tweets related to different disciplines and dissect the sentiments of the tweets. The healthcare assistance sees an increase in requests for emotional analysis in recent times, assaying client feedback(Ruferetal. 2021; Parketal. 2020; Cortis and Davis 2021),[8][7] and client satisfaction analysis are just a few operations. in the field of health care(Baasharetal. 2020; Miottoetal. 2018)[9][10]. The business sector has been using emotion analysis to ameliorate. Emotional analysis of colorful operations similar to quality operation, request exploration, and contender analysis, product review, client feedback, etc. colorful problems are attached with sentimental analysis and NLP. There are many words in different languages whose literals, meaning and environment changes as they're used. Thus, there are not many tools available for all dialects. Sport & affront are two crucial issues that have lately associated experimenters.This study provides an in- depth analysis of sentimental analysis by agitating the field from a variety of perspectives as it covers numerous aspects of exploration related to analysis of sentimental, like operations, tools and styles. This work is of immense importance as it not only suggests the best algorithm

978-1-6654-5319-6/22/$31.00 ©2022 IEEE

for sentiment analysis but also gives a detailed view of the entire process.

III. Procedure

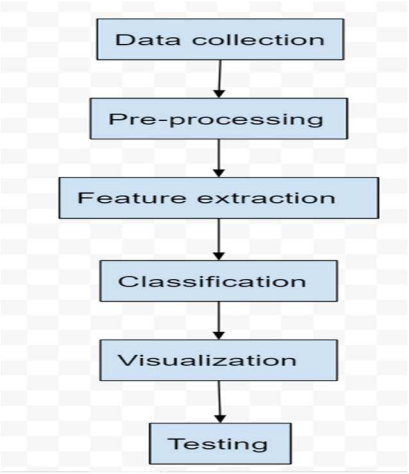

Fig 1. Steps in Opinion Mining

A) Data collection

The very first step is to collect the relevant source of data for Opinion Mining in significant amounts. Dataset used for the same can be extracted through some of the already defined popular datasets like Twitter comments data, Facebook reviews data or posts data. One other form in which data could be collected can be by using a google form for a survey and then analyzing the same data, weblogs, ratings, major discussion forums etc. Direct datasets can also be downloaded from the kaggle platform where datasets are available in humongous amounts.

B) Pre-processing

Pre-processing is the immediate step after cleaning the data where the data is converted to lowercase, all punctuations are removed, segregation of the data, converting them into lists. After collecting the data through forms, or API'S, or different platforms it is highly recommended to make the entire data undergo this step. This step involves removal of the noise data or the unwanted data so that it does not affect the results of pre-processing. Unwanted data may include the numerical data, hyperlinks, special symbols, emails, and words which are highly repetitive and do not contribute to the analysis portion. This step is the foundation step as once the data is cleaned thoroughly and ready for pre-processing it assures even higher degree of both precision and accuracy. There are a number of techniques which are designed for the data cleaning purpose. These methods include range of data, type of data, constraints on data, cross examination of data, are some of the checks required for the data cleaning on the entire dataset.

C) Feature extraction

It's the coming step after pre-processing the entire data. The model to be trained can not be directly trained with the entire data and certain point selection has to be done to reduce the modal complexity and indeed ameliorate delicacy of the model. Certain parameters are to be considered while deciding the different features to be considered while constructing a modal using a standard machine learning algorithm.Feature extraction and feature selection both go hand in hand. Feature extraction is the first subtask where the collection of features to be considered for the model is being minimized.This builds a comprehensive set of features from the original dataset. The goal of feature selection is to create a dictionary in the form of key value pairs where the features are keys and the values are the ranks assigned to each and every feature. This step also removes the set of features which are associated with low ranks

$$Term\ frequency = \frac{number\ of\ x\ terms\ in\ message}{total\ number\ of\ terms\ in\ message}$$

$$Inverse\ Term\ frequency = log(\frac{1}{Term\ frequency})$$

D) Classification

Once the feature is decided from the feature extraction step then it is the time for classification of the data. For Opinion Mining there are mainly 3 categories positive, negative, neutral. Positive Opinion Mining at a token level can include words like 'good', 'happy', 'great', 'excellent', 'Joy', 'Pride', 'Hope', 'Amusement'. Negative Opinion Mining at a token level can include words like 'bad', 'Anger', 'Frustration', 'Fear', 'Guilt', 'Loneliness','Confused','Jealous','Hopeless'. Negative Opinion Mining at a token level can include words like 'Inactive','disinterested','unbiased','Impartial' ,'Indifferent','Fair','Everhanded','Abstract' and many others. After the classification there is a perfect clarity of how many sentences are there with positive, negative and neutral respectively. For a machine to make understand about each and every type of classification present in the data there are several machine learning algorithms which can be used to train on the entire classification.

E) Visualization

After building a machine learning model it's highly recommendable to perform visualization of the classification and the results produced by the machine learning algorithms. For sentinel analysis, on any typical dataset used for training the machine it should be visualized in the form of a graph so that the continuous distribution of the data can be noticed. It clearly means that the data used must contain a variety of sentiments then only the model will predict the test with higher accuracy and precision.

F) Testing

Testing is important to verify the built model as whether it predicts desired output with higher accuracy or not. It can also be called the validation phase. The importance of this step lies in the fact that when it is deployed over large scale applications then it should work properly. Testing is the last phase where a user at run-time provides a machine with a sentence and then the machine predicts about that sentence whether positive, negative, or neutral.

IV. Different Levels of Opinion Mining

A) Token Level

The English literal meaning of the word token is the smallest word that can be formed with letters in english. As the name suggests this level is the basic level where the process of Opinion Mining is entirely done on the basis of each and every single token or a word. For example In the sentence "Although the quality of food was not

that great, I liked the service of the restaurant". The sentence cannot be entirely positive or negative as the very first part of the sentence compiles towards the negative emotion/feedback whereas the second part of the sentence indicates a positive sentiment. Now in this case it may happen that the machine learning algorithm classifies this sentence as a neutral sentence but it can be observed as a human being that the overall response is positive. For such cases we need a token level of Opinion Mining in which we break the entire sentence into subsequent parts, remove unwanted words, preprocess it and then classify the overall sentence as positive,negative or neutral.

B) Sentence Level

This level mainly focuses at the sentence level and is also the primary target from where actually the process of Opinion Mining starts. In English there are different types of sentences particularly declarative, imperative, simple, compound, complex. compound-complex. interrogative. imperative, exclamatory. For example- "He is a good boy." A simple, declarative sentence which clearly indicates positive sentiment. But for a sentence like "I am very disappointed with you because you failed the exam although, you are a good player". Now this is an example of a complex sentence where the model fails to predict the sentiment with high accuracy as it contains more than one emotion. Another example is "The client was happy, though the presentation was not so good, yet he accepted our proposal and signed a contract with us." This is an example of compound-complex sentences where more than two sentences are found. Now in order to correctly identify the sentiment a person can say overall it is a positive sentence because the target was achieved but for a machine to place this type of sentences as a positive, negative or neutral becomes quite challenging when it comes to high degree of precision and accuracy.

C) Paragraph Level

This is the next level after sentence level where the whole paragraph is taken into consideration for calculating sentiments. With the growing trend of social media comments, reviews, feedbacks are a good source to analyze the Opinion Mining at the paragraph level. Nowadays, even captions on a post can be considered for the analysis at the paragraph level.

D) Document Level

This is the last level where Opinion Mining is applicable. In this category the entire document is being classified as a positive, negative or neutral. For example- a detailed description and review of a movie, reviews of books could be some examples where we consider the entire document as a single unit and then try to classify it.

There are different types of machine learning algorithms available to process the data and analyze the result. They can be understood with the help of the following flowchart

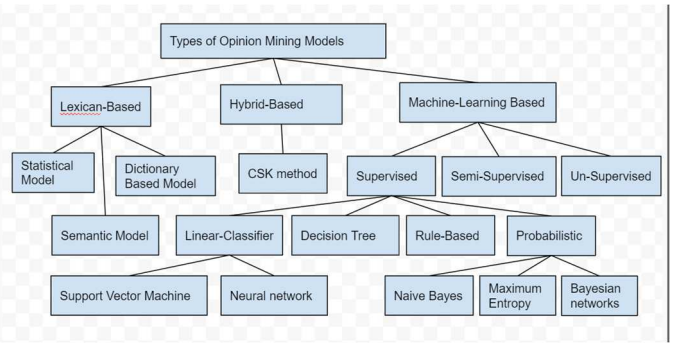

Fig 2. classification of machine learning algorithm

Various Machine Learning models are discussed in this work which.are as follows-

Unsupervised Learning

Unsupervised Learning often comes into use when the training dataset does not contain any labels. For such a dataset it becomes crucial to analyze it properly so that model accuracy does not decrease. Analyzing the data, finding out a relevant pattern and then building a model in coherence with the pattern always gives minimal chance of error. When data is huge, often unsupervised learning is used. Clustering is one such method through which unsupervised machine learning models are trained. Example- When it comes to medical sciences research of symptoms of a disease is done through an unsupervised machine learning model.

Semi-Supervised Learning-

Semi Supervised learning is the hybrid type of machine learning where a part of the data contains a label associated with it whereas the rest of the dataset is completely unorganized. It really minimizes the problems faced in Supervised and Unsupervised learning models and is often used where the dataset is often humungous in size and contains a corresponding label. Clustering is one such method that can be done to build semi-supervised models where data is in humongous amounts and also has it's corresponding labels.

Supervised Learning

In supervised learning often the datasets come with the respective labels or tags. This type of machine learning gives a proper structure to the entire data. Both the training data as well as the testing data is mapped with different labels.Supervised learning makes the task of training a model quite efficiently as the dataset is composed of organized data itself. Example- Weather forecasting where the type of climate is often associated with its temperature can be made using a supervised machine learning model. The different methods in supervised learning are as follows-

i. Support Vector Machine

Support Vector Machine (SVM) and the SVM classifier have been shown to be highly effective in Opinion Mining and generally outperform other algorithms. Support Vector Machine is a new machine learning technique developed based on statistical learning theory and is the most successful implementation of statistical learning theory..

$$W_{tk} = \frac{tft * (log(\frac{A}{Ak}))}{\sqrt{\Sigma\ (tft * (log(\frac{A}{Ak})))^2}} \quad (1)$$

where
W_t = weight of word t in entire dataset
tf_t = frequency of t in doc D,
A = total number of documents in dataset
A_k =documents where word t occurs.

ii. Neural Network

Neural networks often provide an efficient way in building classification systems in which the model learns from the training data as well as the output obtained after each iteration in the model. The whole composition of the neural network is made up of nodes just like the neurons in a human brain. Just like the neurons are used to pass messages from one part of the body to another, nodes in the neural network play a similar role and try to learn from a small amount of data from the training data and build a relationship and classify the data. Neural networks can be broadly classified into two categories:

1. Convolutional neural network-
   Convolutional neural networks are generally used for problems like digital image processing. For example- Hand digit recognition system, Vehicle detection system, Plant disease detection system.
2. Artificial neural network-
   An artificial neural network is related to models where the classification is based on past data and also prediction. For example- Stock value prediction.

iii. Decision Trees

It is a type of data mining induction algorithm that repeatedly segregates data using either a DFS(depth-first search) or BFS(breadth-first search) technique until entire data particulars belong to a certain class. The decision tree structure consists of root, internal and splint bumps. A tree structure is used in the bracket of not known data entries. Each node in the tree, the stylish split is decided using contamination measures. The leaves of the tree correspond to class markers which are divided into data particulars. The decision making tree bracket fashion is done in two stages: trees' structure and trees' pruning. The construction of the tree is done from top to bottom. Recursively partitioned is done in this stage until all data particulars correspond to the same class marker.

Given a collection or sets of data A(A1,A2,.....An). Consider there are i distinct class labels defining the i different class values in B (1….. i). Let A_o be the total number of samples in A in class B_i. Following calculations are performed to extract information from data and classify it

$$Split\_Info_A(A) = -\sum((|A_o|\backslash |A|) * \log(|A_o| \backslash |A|)) \quad (1)$$

$$Gain\ ratio(A) = Gain(A)/ Split\_info_A(A) \quad (2)$$

where,

$$Gain(A) = Info(A) - Info_A(A) \quad (3)$$

$$Info(A) = -\Sigma P_i * \log_2(Pr_i) \quad (4)$$

and

$$Info_A(D) = -\Sigma (|A_o| / |A|) * Info(A_o) \quad (5)$$

where,
Pr_i = probability of distinct class B_i,
A =data set,
O=sub-attribute from attribute,
($|A_o| / |A|$)= weight of o^{th} partition.

These formulas are helpful in classifying each attribute according to the set and then making the classification of the model.

iv. Naive Bayes

Naive Bayes algorithm uses the concept of probability at the token level in order to calculate and efficiently classify data as a positive or a negative sentiment. The whole dataset is divided into sentences followed by individual tokens after which polarity of the tokens decide the overall sentiment of the sentence and classifies it into positive, negative or a neutral statement.

$$P(a|b1, b2,..., bn) = \frac{P(a)P(b1,b2,...,bn|a)}{P(b1,b2,...,bn)} \quad (1)$$

where,
$P(a|b_1, b_2,..., b_n)$ is probability for all successive tokens.
P(a) is cumulative class probability.
$P(b_1, b_2,..., b_n)$ is the prior probability of feature set.

In a general manner it can be the above formula can be written as

$$P(b_1, b_2,..., b_n|a) = \prod_{i=1}^{i=n} P(b_i|a) \quad (2)$$

v. Maximum Cross entropy

This algorithm also supports a probabilistic approach like the Naive Bayes algorithm. The main difference is to maintain a set of user extracted features designed and accordingly weights are calculated by the model. It's main motive is to maximize the entropy and according to the extracted features in sync with the constraints specified by the user. In order to maximize entropy it is only achieved when all events are mutually expected to happen and the happening of a single event doesn't affect the probability of others. Its goal is building an unbiased model so that at the time of classification probability distribution remains uniform.
Assuming only the word level features, we declare a joint function f(w, c) = N, for every word w and class cC, where, N is frequency of w occurs in a query in class C. while optimizing it iteratively, we assign a weight to each joint feature to maximize the log-likelihood of the training data. The probability of class c given a query q and weights λ is

$$P(c|q, \lambda) = \frac{exp\Sigma\ i\ \lambda i f i\ (c,q)}{\Sigma\ c\ 0\ C\ exp\Sigma\ i\ \lambda i f i\ (c\ 0\ ,q)} \quad (1)$$

TABLE I. MAXIMUM CROSS ENTROPY CLASSIFIERS

vi. K-NN

K-NN works similarly to Naive Bayes algorithm. The main difference lies on the basis of selecting the feature to be chosen as K-NN uses the cosine or the Euclidean distance in order to calculate the nearest neighbor for any node K. The euclidean distance between K and any of it's neighbor can be calculated as follows-

$$D(t,z)=(\sum_{r=1}^{d} |xr-zr|^p)^{1/p} \quad (1)$$

where,

D(t,z) = distance

t =a test input

z = reference input

d = total number of such neighbors

P = 1,2,3… ∞

It considers the fact that the objects closer in distance are mostly similar in nature and therefore impacts from its subsequent neighbors. In parallel to distance, a collection of weights W, for every closest neighbor is maintained.

V. Related Work

Multiple publications can be found in the domain of opinion mining which uses a suitable supervised learning machine algorithm to build a model and classify data with best of their respective accuracies. A lot of work has been done in different languages like English, Chinese, Urdu in the same field.Supervised Machine Learning Algorithms provide high accuracy to the problem described in this paper and also classifies text with high precision. Our main motive is to dig deep in the topic Opinion mining in Indian Languages for which there is a considerable amount of research being done on the major language like hindi. The author in [3] used SVM to build a model using a hotel review dataset in the hindi language.The author in [4] used newspaper articles, blogs, websites in hindi to construct a dataset and then tried building a model using Supervised Machine Learning algorithm for the first time for Hindi dataset. The authors in [2][5][8][12][13] were the recent ones which used Twitter dataset for their respective analysis and used different machine learning algorithms and even gained 79.9%, 84.6%, 77%, 78%, 90% accuracy respectively. Indian Languages have still remained untouched on this topic as major research has not been in any Indian Language.

[1] MACHINE LEARNING MODELS

S.NO	Machine Learning Models		
	Research paper title	***Dataset***	***Accuracy***
1	Affective-feature-based Opinion Mining using SVM Classifier [11]	ChnSentiCorp corpus	73.9%
2	Opinion Mining of healthcare Tweets using SVM Classifier [12]	Twitter API (twitter4j)	79.9%
3	Sentimental Analysis using Support Vector Machine[13]	Hotel reviews dataset	69.7%
4	Aspect based Sentiment Analysis in Hindi: Resource Creation and Evaluation[14]	Hindi newspapers, blogs	61.9%
5	Opinion Mining using Neural Networks: A New Approach [15]	Twitter Dataset	84.6%
6	Opinion Mining and Prediction using Neural Networks [16]	Facebook dataset	84.0%
7	Opinion Mining for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning [17]	Hotel reviews dataset	90.9%
8	Opinion Mining about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes[18]	Twitter review dataset of Indonesia	77%
9	Sentimental Analysis using Naive Bayes Classifier[19]	Amazon product review dataset	89%
10	Machine Learning Techniques for Sentiment Analysis of Indian Languages [20]	Indian Language dataset	75%
11	A Naive Bayes and Maximum Entropy approach to Opinion Mining: Capturing Domain-Specific Data in Weibo[21]	Chinese social media	90.6%
12	Opinion Mining about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes[22]	Twitter review dataset of Indonesia	78%
13	Opinion Mining of Law Enforcement Performance Using Support Vector Machine and K-Nearest Neighbor [23]	Twitter	90.2%

VI. Conclusion

This work includes all the different types of classification models that have been discussed along with their respective accuracy and algorithms. The type of dataset used often decides the type of machine learning algorithm required to train the machine. It is observed that Supervised machine learning provides the ultimate solution with higher accuracy and precision when building linear classifiers at the time of training the model as the data is well characterized and often associated with labels. With the different supervised algorithms being discussed namely SVM, Neural Network, Decision Trees (DT), Naive Bayes, Maximum Cross entropy, K-NN the DT algorithm provides maximum accuracy and efficiency with highest precision of almost 90% approximately.

REFERENCES

[1] Subhashini LD, Li Y, Zhang J, Atukorale AS, Wu Y. Mining and classifying customer reviews: a survey. Artificial Intelligence Review. 2021 Dec;54(8):6343-89...

[2] Mowlaei ME, Abadeh MS, Keshavarz H. Aspect-based sentiment analysis using adaptive aspect-based lexicons. Expert Systems with Applications. 2020 Jun 15;148:113234.

[3] Prasad SS, Kumar J, Prabhakar DK, Pal S. Sentiment classification: an approach for Indian language tweets using decision tree. InInternational conference on mining intelligence and knowledge exploration 2015 Dec 9 (pp. 656-663). Springer, Cham..

[4] Zvarevashe K, Olugbara OO. A framework for sentiment analysis with opinion mining of hotel reviews. In2018 Conference on information communications technology and society (ICTAS) 2018 Mar 8 (pp. 1-4). IEEE..

[5] Valencia F, Gómez-Espinosa A, Valdés-Aguirre B. Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. Entropy. 2019 Jun 14;21(6):589..

[6] Ahmad S, Asghar MZ, Alotaibi FM, Awan I. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Human-centric Computing and Information Sciences. 2019 Dec;9(1):1-23.

[7] Cortis K, Davis B. Over a decade of social opinion mining: a systematic review. Artificial intelligence review. 2021 Oct;54(7):4873-965.

[8] Perrin J, Streeck N, Naef R, Rufer M, Peng-Keller S, Rettke H. Comparing perspectives: patients' and health care professionals' views on spiritual concerns and needs in chronic pain care-a qualitative study. BMC Health Services Research. 2021 Dec;21(1):1-1.

[9] Baashar Y, Alhussian H, Patel A, Alkawsi G, Alzahrani AI, Alfarraj O, Hayder G. Customer relationship management systems (CRMS) in the healthcare environment: A systematic literature review. Computer Standards & Interfaces. 2020 Aug 1;71:103442.

[10] Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics. 2018 Nov;19(6):1236-46.

[11] Luo F, Li C, Cao Z. Affective-feature-based sentiment analysis using SVM classifier. In2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD) 2016 May 4 (pp. 276-281). IEEE.

[12] Saranya G, Geetha G, Meenakshi K, Karpagaselvi S. Sentiment analysis of healthcare Tweets using SVM Classifier. In2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS) 2020 Dec 10 (pp. 1-3). IEEE.

[13] Zainuddin N, Selamat A. Sentiment analysis using support vector machine. In2014 international conference on computer, communications, and control technology (I4CT) 2014 Sep 2 (pp. 333-337). IEEE.

[14] Akhtar MS, Ekbal A, Bhattacharyya P. Aspect based sentiment analysis in Hindi: resource creation and evaluation. InProceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) 2016 May (pp. 2703-2709).

[15] Dhar S, Pednekar S, Borad K, Save A. Sentiment analysis using neural networks: A new approach. In2018 Second international conference on inventive communication and computational technologies (ICICCT) 2018 Apr 20 (pp. 1220-1224). IEEE.

[16] Paliwal S, Khatri SK, Sharma M. Sentiment analysis and prediction using neural networks. InInternational Conference on Advanced Informatics for Computing Research 2018 Jul 14 (pp. 458-470). Springer, Singapore.

[17] Yang L, Li Y, Wang J, Sherratt RS. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. IEEE access. 2020 Jan 27;8:23522-30..

[18] Bayhaqy A, Sfenrianto S, Nainggolan K, Kaburuan ER. Sentiment analysis about E-commerce from tweets using decision tree, K-nearest neighbor, and naïve bayes. In2018 international conference on orange technologies (ICOT) 2018 Oct 23 (pp. 1-6). IEEE.

[19] Surya PP, Subbulakshmi B. Sentimental analysis using Naive Bayes classifier. In2019 International conference on vision towards emerging trends in communication and networking (ViTECoN) 2019 Mar 30 (pp. 1-5). IEEE.

[20] Ahmad GI, Singla J. Machine learning techniques for sentiment analysis of indian languages". Int. J. Recent Technol. Eng.;8(2):3630-6.

[21] Ficamos P, Liu Y, Chen W. A Naive Bayes and Maximum Entropy approach to sentiment analysis: Capturing domain-specific data in Weibo. In2017 IEEE International Conference on Big Data and Smart Computing (BigComp) 2017 Feb 13 (pp. 336-339). IEEE.

[22] Bayhaqy A, Sfenrianto S, Nainggolan K, Kaburuan ER. Sentiment analysis about E-commerce from tweets using decision tree, K-nearest neighbor, and naïve bayes. In2018 international conference on orange technologies (ICOT) 2018 Oct 23 (pp. 1-6). IEEE.

[23] Istia SS, Purnomo HD. Sentiment analysis of law enforcement performance using support vector machine and K-nearest neighbor. In2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE) 2018 Nov 13 (pp. 84-89). IEEE.