

# Sentiment Analysis of X(Twitter) Data: A Review Study

**Arshpreet Kaur**

Assistant Professor, Computer Science

## **Abstract:**

Social networking Sites have captured the attention of the every generation. The advancement of technology and social networking sites has enabled anyone to easily express and share their ideas, opinions, feelings, expressions and reviews about any topic like- social, political, brands, business etc, with other people around the world. Thus Social media contains large amount of data in the form of text, image, audio and video uploaded by users. This data can be used as valuable information to check the reviews of people. It is also possible to diagnose every sentiment of the user by his/her text and posts on social media using emotion theories, machine learning techniques, and natural language processing techniques. This paper reviews the recent studies of sentiment analysis by social networking site X(Twitter). The steps of the methodology and related work analysis of the existing system is discussed in this review paper.

**Keyword:** Social Media,X(Twitter), Machine Learning

## **1. Introduction**

The use of social networks has rapidly grown in the previous years. Internet and social sites has become basic requirement in this social media era. . According to the statistics from United Nations, by February of 2019, the current worldwide population is 7.7 billion. The internet has 4.2 billion users, which is more than half of the population in the world. Social media takes the greatest number of users with 3.397 billion active social media users, which is 44% of the worldwide population (Statista, 2020)[1]. Social media has had a huge and great effect on communication and connection of lives of human beings.. People express their opinions,views and reactions about political,social,brand,business etc. There are various social media platforms like Facebook, X(Twitter), Wikipedia, LinkedIn, Google+, Instagram etc.In these social sites people express their views in the form of text,image,audio and video.Its great for users of social networking sites to freely and openly contribute and communicate online about any topic. So its easy to gather and analyze peoples' reactions towards any topic.It is impossible to manually gather and analyze the opinions of people on social networks. Therefore several techniques are used to analyze the sentiments of people About any topic and then check the polarity of the opinion as positive, negative or neutral. Machine learning techniques based on natural language processing ,lexicon technique and deep learning can be used for this purpose. It is possible to analyze the sentiments by text data and image data. This paper is divided in six section as –Introduction, Basics of X(twitter) and sentimental analysis, Approaches for sentimental analysis, General Steps for Twitter Sentiment Analysis in Machine Learning Approach , Analysis of Related Work and conclusion.

## 2. X(Twitter) and Sentiment Analysis

### 2.1 X(Twitter)

X(Twitter) is a open micro-blogging social network that people use to communicate with each other and express themselves in the form of short messages, Known as Posts(tweets). Posts are basically restricted to 280 Characters . X(Twitter) has millions of users who share their views by Posts(tweets). It has 240+ million active users. About 500 million posts(tweets) are generated each day [2]. In X(twitter) you follow people and other people follows you. When you share anything, your followers can see it. You can also repost(retweet) something by mentioning them with @ sign ,so that your followers can see the post of that particular person. You can also do messages privately, if you do not want to share anything publically. You can also put phrases to in your posts(tweets) with #(hashtag).X is mainly used for sentiment analysis because X has variety of users and posts are small in length and thus easy to analyze.X(Twitter) is most commonly site used for sentiment analysis.

### 2.2 Sentiment Analysis

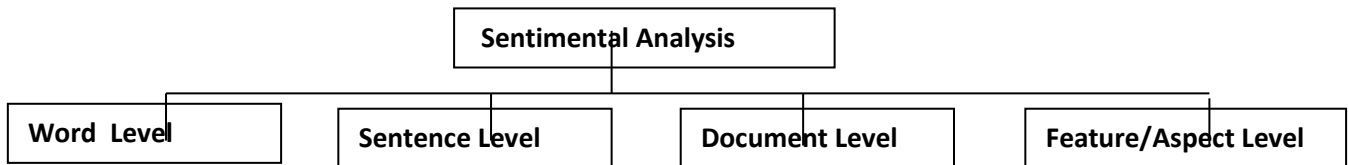
Sentiment Analysis is data-mining technique, it is also known as text-mining or opinion mining. Sentimental Analysis is used to analyze the online expressions. Sentiment analysis (SA), is a process that applies natural language processing(NLP) text analysis and machine learning to identify the sentiment of the text as positive, negative or neutral.Thus machines are trained with texts inputs and machines take further decisions without human intervention. There are two types of sentiment classification techniques, binary classification technique and multi-class sentiment classification technique. In binary classification output classified in two categories as (Happy or Unhappy) and in multi-classification output is classified as (Strongly happy,Very happy,less happy,Unhappy etc). Previous researches show that Sentiment Analysis is domain centred, i.e. results of one domain cannot be applied to other domain [3].So the performances of sentiment classifiers are dependent on topics. Because of that we cannot say that one classifier is the best for all topics since one classifier doesnt consistently outperforms the other. Sentiment Analysis is one of the most popular application in text mining and with the integration of machine learning algorithms, deep learning algorithms it becomes more effective and is used in large number of businesses to increase their productivity and to make better customer experience. Sentiment Analysis in X is quite difficult due to its short length, posts are generally written in informal language,use of acronyms and abbreviations, Presence of emoticons, slang words and misspellings in tweets forced to have a preprocessing step before feature extraction.

#### 2.2.1 Types of sentimental analysis

1. **Fine Grained sentiment analysis:-** It is a binary sentiment analysis which determine the polarity of texts as positive/negative or depressed/non-depressed.It can be muti-specification type as (Strongly happy, very happy ,less happy, unhappy).
2. **Emotion detection sentimental analysis:-**It is sophisticated way to identify signs of emotional states from piece of text( sad, angry, happy, worry etc).Laxicons and machine learning are complex algorithmic methods are used to determine the sentiments.
3. **Aspect based sentimental analysis:-** It is deep study of sentiments. It identify and represents the opinion regarding element. It analyze the different aspects of the sentence or tweet.
4. **Intent Analysis :-** It's purpose is to identify intentions expressed in any text(interested, not interested ).It is deeper understanding of the intentions.

### 2.2.3 Levels Of Sentiment Analysis

Depending upon the requirements sentimental analysis algorithms can be used at the following scopes.



**Fig-1 –(Sentimental Analysis Levels)**

- A. Word-level sentiment classification:-** Word level sentimental analysis checks the sentiments of words as positive , negative or neutral.
- B. Sentence-level sentiment classification:** It obtains the sentiment of a single sentence. At this level, each sentence is classified as “positive”, “negative”.
- C. Document-level sentiment classification:-** It is analysis of the entire document. At this level, a document can be classified entirely as “positive”, “negative”, or “neutral”.
- D. Feature/Abstract Level :-** It is deep level of sentimental analysis. It analyze the different aspects of the sentence or tweet. E.g. – The phone quality is good but battery life is disappointing. In this example aspects are quality and battery.Aspect level sentiment analysis is superior to other levels.

### 3. Approaches of Sentimental Analysis

There are different approaches to perform sentiment analysis like-Machine learning approaches, Lexical based approaches and Hybrid approaches.

#### 1. Machine learning approach:-

In this machine learning based algorithms used to classify the given data. It can further divided as Supervised learning and Unsupervised learning. Supervised machine learning has a training set along with a testing set. Firstly machine is trained by providing the enough amount of data and then testing is performed to check the working. Following are the Supervised ML algorithms-SVM, Logistic Regression, Naive Bayes etc. Unsupervised machine learning does not contain a training set.In this unlabeled data is provided and the algorithm classifies the data based on patterns. The various algorithms of unsupervised ML are- K-Means, DB-Scan clustering etc.

#### 2. Lexical approach:-

This technique uses large number of words to and classify the polarity of texts. It is classified into dictionary and corpus based approaches. Dictionary based approaches utilize a dictionary to analyze the polarity of a sentence. Its positive point is that it does not require any training data. But its weak point is that many words are not included in sentiment lexicons.

#### 3. Hybrid Approach:-

It is combination of machine learning and lexicon-based approaches to address Sentiment Analysis.

#### 4. Ensemble method –

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model . Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would. Ensemble method is also used for sentiment analysis by many researchers.

### 5. Deep Learning-

Deep Learning is a Machine Learning methods class. Deep Learning uses artificial neural networks to perform sophisticated computations on large amount of data. It is a type of machine learning that works based on structure and functions of human brain. It extracts high features progressively by using multiple layers. The arrival of deep learning techniques has opened new doors for fresh possibilities and horizons. Deep understanding requires deep neural networks in learning multifaceted features. These features have been extracted from the assistance of minimum external contributions. The deep learning approach required a large dataset to get a significant boost in the performance of a model. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers. The various algorithms of deep learning are Convolutional Neural Network (CNNs), Long short Term Memory Network (LSTMs), Recurrent Neural Network (RNNs), Multilayer Perceptrons (MLPs), Deep Belief Networks (DBNs) etc.

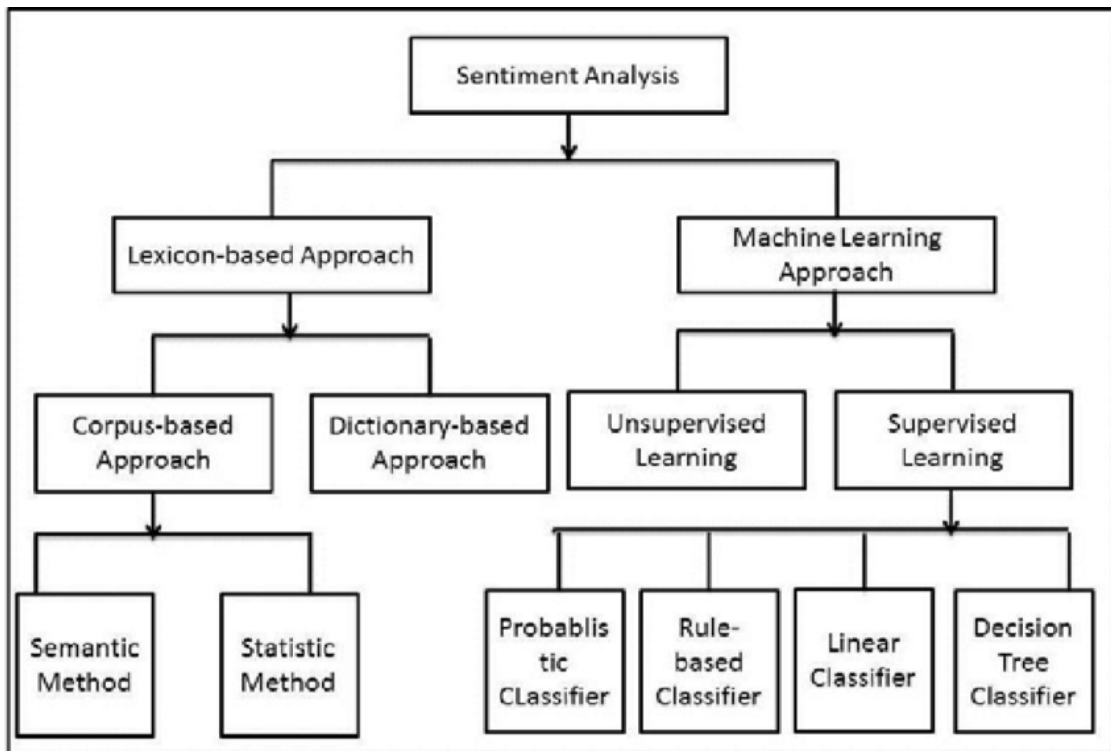


Fig-2 ( Approaches for Sentimental Analysis [4] )

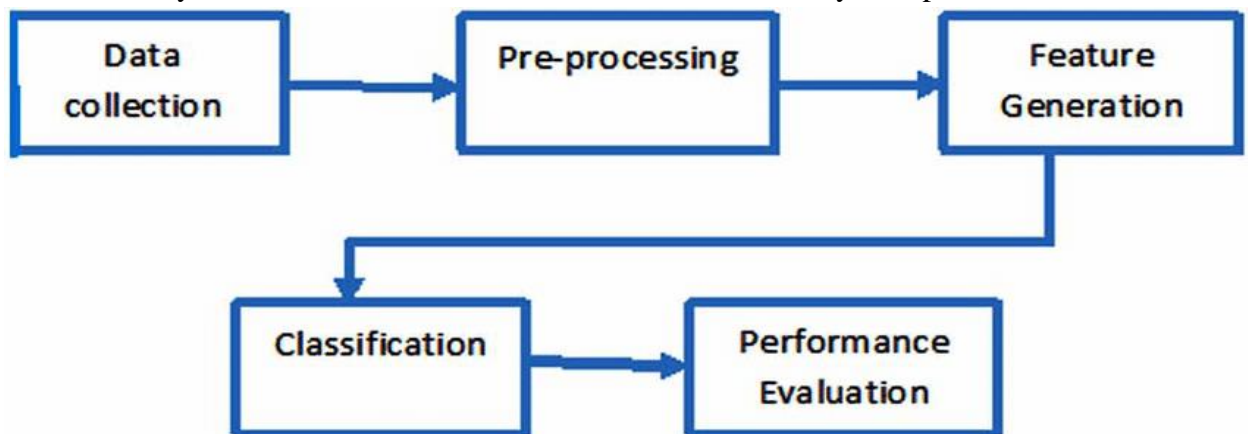
### 4. General Steps for X(Twitter) Sentiment Analysis In Machine Learning Approach

1. **Data Collection:-** Data collection from x(Twitter) API or user profiles.
2. **Pre-Processing:** It is the process to remove extra features from the data. It will decrease the size of the Post by removing extra records. As removal of special symbols, punctuations, numbers, stop words URLs etc.
3. **Feature Extraction:-** It is the process of extracting features from the posts. It will Consider following features –

- A. **Parts of Speech (POS)**-Analyzing the various parts of speech as verb,adverb,adjective etc.Then analyzing that in what sense these words are used and sentiment analysis is performed.
- B. **Words & word frequencies**- It is co-occurring of words represented as Uni-gram,Bi-grams and N-grams
- C. **Position Of Terms**-The position of term can change the overall sentiment of the text.
- D. **Opinion Words and Phrases**-Phrases or idioms can be used as feature. To check sentiment of text.
- E. **Negation**-Negation usually change the polarity of the opinion i.e. not good, not bad.

**4. Classification:** This step is responsible for classification of posts. Whether the post expresses a depressed or non-depressed sentiment. There are three techniques to classify the sentiment of a post they are, machine learning approach, lexicon based approach and hybrid approach. The results of all these methods gives different accuracy level.

**5. Performance Evaluation:** This step is useful in determining the accuracy of the particular classifier. Performance is usually determined in terms of recall, f-measure , accuracy and precision.



**Fig-3 (Steps in X(Twitter) sentiment analysis [2] )**

### 5. Analysis of Related Work

In this Era of social networking, Sentimental analysis is well established globally. Many work has been done on this. The below given table shows the work accomplished by different authors on X(Twitter) sentimental analysis. The various approaches and accuracy level of different works is given in the table.

S.No.	Reference	Title	Method Used	Accuracy
1	[5]	Sentiment Analysis of Tweets on Social Issues using Machine Learning Approach	Supervised Machine learning	NaïveBayes-85.49%, SVM-89.91%, Decision tree 86.14%, random forest 88.36%, K Neighbors 75.32 %, logistic regression 86.63% SGDC 90.57%.
2	[6]	Deep Learning for Automated Sentiment	Deep Learning	LSTM 80.83% BiLSTM 87.17%

		Analysis of Social Media		GRU 64.92%
3	[7]	Depression Detection on Social Media Network (Twitter) using SA	Supervised classification techniques	87.5% with linear classifier.
4	[8]	Twitter Sentiment Analysis using Deep Learning Methods	Deep Learning	MLP (Multilayer Perceptron)- 67.45% Deep Learning Methods- 77.45 %
5	[9]	Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression	Supervised classification techniques	SVM-91% Naïve bayes-83% Maximum Entropy 80 %.
6	[10]	Sentiment Analysis on Twitter Data Using Deep Learning Approach	Deep Learning	SGDClassifier-65.87% LogisticRegression-68.37%, LogisticRegressionCV 68.37%, LinearSVC( SVM)-65.80% , RandomForestClassifier -64.62 %, CNN-78.81%
7	[11]	Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning	Supervised ML	The SVM using the hybrid feature selection achieved an accuracy of 88%. In addition, the SVM with the PCA achieved an accuracy of 92%.
8	[12]	<i>Social Media Sentiment Analysis On Twitter Datasets</i>	Supervised ML	Decision tree-99.3% SVM-91.6% Tandom forest-99.4%
9	[13]	A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques	Supervised ML	Logistic regression-86.23 % SVM-85.69 % Multinomial naïve bayes - 83.54%
10.	[14]	Twitter Sentiments Analysis Using Machine Learning Methods	Supervised ML	Naïve Bayes Classifier- 86 % Support Vector Machine-74.6% Maximum Entropy Method -82.6 %
11	[15]	Evolutionary Ensemble of Sentiment Analysis Methods for Domain Adaptation	Ensemble	The ensemble method achieved a macro F1-score of 62.98%. However, the ranking algorithm and skip-gram obtained a macro F1 score of 61.60%
12	[16]	Simple approaches of sentiment analysis via ensemble learning	Ensemble	The ensemble classifier obtained an Fscore of 86.05%
13	[17]	Gurmukhi text emotion	Supervised	Decision Tree shows high values of

		classification system using TF-IDF and N-gram feature set reduced using APSO	Machine learning	recall, F-score and accuracy in range 0.82 to 1.
--	--	--	------------------	--

## 6. Conclusion

This review paper gives the basic information about sentimental analysis of X(Twitter) data. This paper discusses various levels of sentimental analysis, approaches, steps in case of machine learning. The comparative study of various works gives us the idea about the efficiency of various approaches. It represents that various approaches can be used for sentiment analysis with different accuracy level. There are different approaches for sentimental analysis. Machine learning and Deep Learning algorithms can be compared and can be ensemble to get best results.

## References

1. Z. Wang and Z. Wang, "SENTIMENT ANALYSIS OF CLIMATE CHANGE USING TWITTER API AND MACHINE LEARNING," 2020.
2. I. Nazeer, M. Rashid, S. K. Gupta, and A. Kumar, "Use of Novel Ensemble Machine Learning Approach for Social Media Sentiment Analysis," no. January 2021, pp. 16–28, 2020, doi: 10.4018/978-1-7998-4718-2.ch002.
3. R. S. Wadawadagi and V. B. Pagi, "Sentiment Analysis on Social Media," no. April, pp. 508–527, 2020, doi: 10.4018/978-1-5225-9643-1.ch024.
4. K. K. Pawar, P. P. Shrishrimal, and R. R. Deshmukh, "Twitter Sentiment Analysis: A Review," vol. 6, no. 4, pp. 957–964, 2015.
5. C. Kaur, "Sentiment Analysis of Tweets on Social Issues using Machine Learning Approach," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 6303–6311, 2020, doi: 10.30534/ijatcse/2020/310942020.
6. L. Cheng, "Deep Learning for Automated Sentiment Analysis of Social Media," pp. 1001–1004, 2019, doi: 10.1145/3341161.3344821.
7. P. S. J. Pachouly, G. Raut, K. Bute, R. Tambe, and S. Bhavsar, "Depression Detection on Social Media Network ( Twitter ) using Sentiment Analysis," pp. 1834–1839, 2021.
8. A. M. Ramadhani, "Twitter Sentiment Analysis using Deep Learning Methods," 2017.
9. A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," *Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Technol. Lead. Fourth Ind. Revolution, ICTC 2017*, vol. 2017-Decem, pp. 138–140, 2017, doi: 10.1109/ICTC.2017.8190959.
10. V. Tyagi, "Sentiment Analysis on Twitter Data Using Deep Learning approach," pp. 187–190, 2020.
11. IEEE Communications Society, Association for Computing Machinery. Special Interest Group on Data Communications, and ACM SIGMOBILE, "2014 Sixth International Conference on Communication Systems and Networks (COMSNETS): January 7-10, 2014, Bangalore, India," 2014.
12. S. Tiwari, "Social Media Sentiment Analysis On Twitter Datasets," pp. 925–927, 2020.
13. A. Poornima, "A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques," pp. 493–496, 2020, doi: 10.1109/ICACCS48705.2020.9074312.

14. L. Mandloi, “Twitter Sentiments Analysis Using Machine Learning Methods,” pp. 1–5, 2020.
15. M. López, A. Valdivia, E. Martínez-Cámara, M. V. Luzón, and F. Herrera, “E2SAM: Evolutionary ensemble of sentiment analysis methods for domain adaptation,” *Inf. Sci. (Ny)*, vol. 480, pp. 273–286, 2019, doi: 10.1016/j.ins.2018.12.038.
16. S. Saadah and G. S. Wulandari, “Information Science and Applications,” *Lect. Notes Electr. Eng.*, vol. 339, no. iii, pp. 777–784, 2015, doi: 10.1007/978-3-662-46578-3.
17. R. Kaur and V. Bhardwaj, “Gurmukhi text emotion classification system using TF-IDF and N-gram feature set reduced using APSO,” *Int. J. Emerg. Technol.*, vol. 10, no. 3, pp. 352–362, 2019.