

Machine Learning-Based Sentiment Analysis in Social Networks: A Survey Review

Uzara Khan, Mahesh Motwani, Anjna Jayant Deen

Department of Computer Science and Engineering, University of Information Technology, RGPV, Bhopal, 462033, India

Abstract

Sentiment Analysis is the process of detecting and categorizing opinions expressed in a piece of text. Since it deals with opinions, thus it is also known as Opinion Mining. It is used to determine whether the writer's attitude towards any particular topic, product, etc is positive, negative, or neutral. It is also called Emotions Extraction. Since a huge amount of data is getting produced on by ventures, Machine Learning technique has become essential to discover business intelligence from data for deliberate decision. The main objective behind this is that people of all walks of life share their ideas and opinions/sentiments over social media platforms like Twitter, Facebook, YouTube, LinkedIn, Instagram, etc. Hence, there exists social feedback in online social network. Collection of business-related data from social media and by using machine learning algorithms to extract useful opinions from them bestows a competitive edge to the enterprise. A survey about various sentiment analysis techniques, challenges, and applications has been covered in this paper.

Keywords: Sentiment Analysis, Opinion Mining, Literature Review, Sentiment Analysis Techniques, Sentiment Analysis Applications and Challenges.

1. INTRODUCTION

1.1 Sentiment Analysis:

Sentiment Analysis is the process of influential whether the quantity of text is positive, negative, or neutral. Sentiment analysis for text combines natural language processing and machine learning techniques to assign weighted sentiments scores to the ideas, topics, themes, etc. It allows a person to understand emotions and opinions. The objective of sentiment analysis is to determine the thought or attitude of a speaker or a writer. Sentiment Analysis is not only limited to text, it can be applied to images, audio, and video as well. Today technology has become advanced because of the appetizer of the internet[7]. Today everyone can express their emotions using online blogging sites or social networking platforms. Therefore, the mining of these data and extraction of sentiment has become an important field of research.

Sentiment analysis and opinion mining reproduce nearly the same meaning. Sentiment analysis recognizes emotion expresses in a natural language text and opinion mining is active to extract the opinion from the text.

1.2 Subjectivity/Objectivity:

To perform sentiment analysis, we first need to identify the subjective and objective text. The only subjective text holds the sentiments[5]. The objective text contains only factual information. Example-

- **Subjective:** Ramayana is a wonderful television show
(this sentence has a sentiment (wonderful); thus, it is subjective).
- **Objective:** Ramanand Sagar is the director of Ramayana.
(this sentence has no sentiment, it is a fact, thus it is objective).

1.3 Polarity Classification:

Polarity classification is the task that distinguishes sentences that express positive, negative, or neutral polarities.

- Positive- My experience so far has been fantastic/brilliant.
- Negative- Dissatisfied with the service.
- Neutral- I will expect a lot more in the future

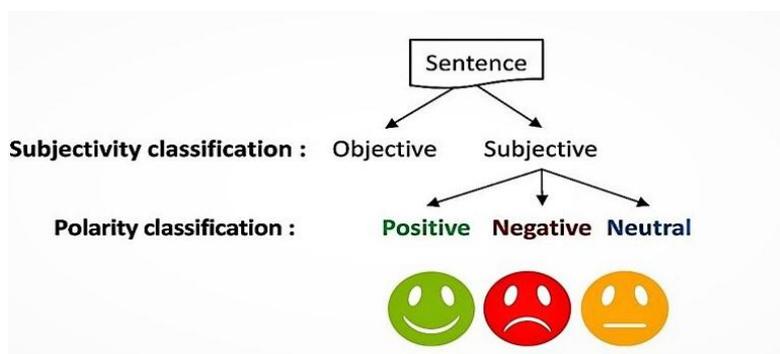


Figure 1: Sentiment analysis workflow.

1.4 Sentiment level:

sentiment analysis can be performed at several levels –

- **Document Level-** In it, the entire document is given a solitary polarity positive, negative, or neutral[13].
- **Sentence Level-** In it, the document is classified at the sentence level. Each sentence is studied separately and classified as sentiments. Thus, the complete document has many sentences where each sentence has its polarity[21].
- **Phrase Level-** It involves a much a deeper analysis of text and deals with the identification of the phrases or aspects in a sentence and analysing the phrases and categorize them as sentiments[17].

1.5 Social Network Analytics Impact on Sentiment Analysis

The exponential development in the usage of digital devices, together with a ubiquitous online contact affords extraordinary ground for the constant connectivity of people and offers marvellous capabilities for publicly expressing opinions, attitudes, or reactions regarding many aspects of everyday human activities. Social media, such as blogs, forums, and social network platforms are a very integral part of people's lives, the virtual spaces where daily individuals share opinions and information and maintain or expand their relational network[15]. we believe businesses shouldn't have to sacrifice speed, scale, or accuracy to understand what consumers are saying about their products. Our opinion mining and sentiment analysis solutions combine AI with real human insight to deliver the most accurate results, helping product reviews, and

other's opinions complete large- scale sentiment analysis projects in days. The massive use of online social networks and the abundance of data collected through them has raised exponentially the attention of the scientific and business community toward them.

Coming from handling the complex characteristics of natural language without considering the data collected through social networks as networked data. Most of the work in sentiment analysis is based merely on textual information expressed in online posts and comments. Early approaches to overcome this important limitation are emerging in recent literature, trying for example, to leverage information on friendship relationships between individuals, since connected users may be likelier to hold similar opinions[21]. However, these features only approximate the rich relation structure encoded in an online social network.

1.6 Online Social Network Definitions And History:

Social networking is the use of Internet-based social networking sites to stay connected with friends, family, colleagues, consumers, or clients. SixDegrees.com had more than 1 million users and was active until 2001. Starting in the first decade of this century, many social networking sites were born, trying to capitalize on the winning ideas of SixDegrees.com. In these years, popular names, including Friendster, Myspace, Facebook, and YouTube, LinkedIn, Instagram[23].

Social networking can have a social purpose, a professional purpose, or both, through sites such as Facebook, Twitter, LinkedIn, and Instagram, among others. Social networking has become a significant root for marketers seeking to engage customers. Market spearhead Facebook was the first social network to beat one billion registered accounts and currently assertions around 2.5 billion monthly active users, making it the most popular social network worldwide[19]. By comparison, the photo-sharing app Instagram had one billion monthly active accounts.

In June 2019, the most downloaded social networking apps in the Apple App Store included mobile messaging apps WhatsApp and Facebook Messenger, as well as the ever-popular app version of Facebook.

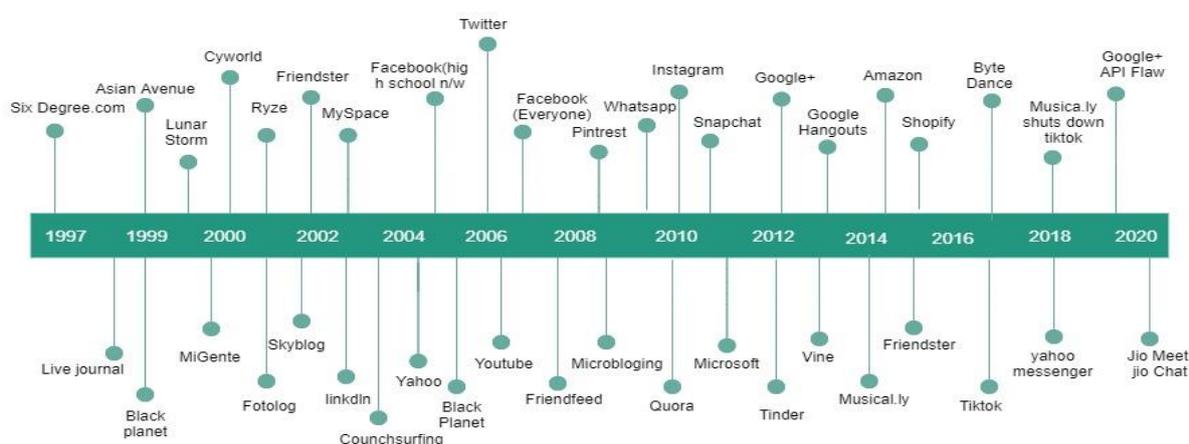


Figure 2: History of Social Network.

There are 7.7 billion people in the world, with at least 3.8 billion of us online. This means social media platforms are used by one-in-three persons in the world, and further than two-thirds of all internet users. Social media has changed the world[9]. The quick and huge adoption of these

technologies is changing how we find buddies, how we access information from the news, and how we organize to demand politically aware changes.

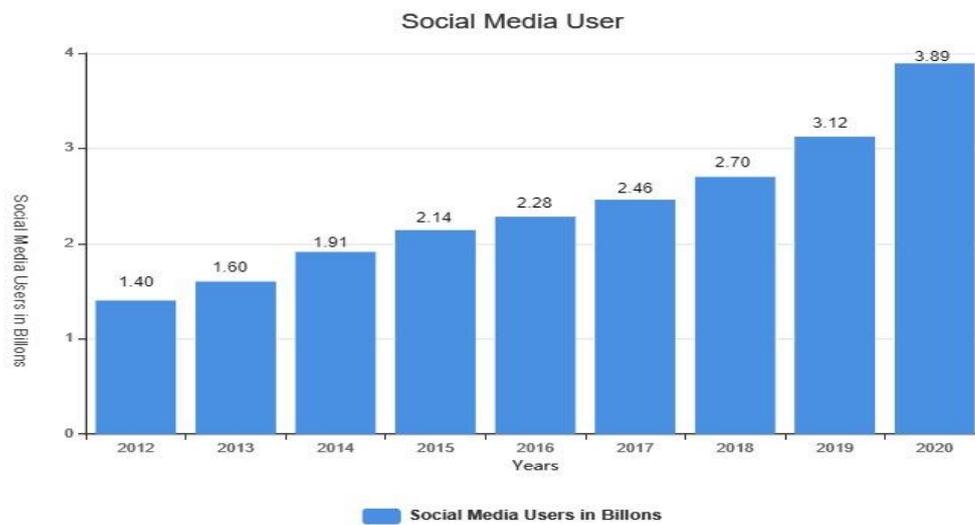


Figure 3: Social Media Users.

Now 2020 the regularly uses social networks are Instagram, Facebook, YouTube, we begin with an outline of key trends and conclude with a perspective on the rate of adoption of social media relative to other modern communication technologies[10].

1.7 Stages Sentiment analysis:

Review: Before you can use a sentiment, analysis model needs to find the product reviews you want to analysed implement a procedure for review and approval of data.

Data Preparation: The data preparation steps complete necessary data pre-processing and cleaning on the dataset for the subsequent analysis. Some commonly used pre-processing steps include removing non-textual contents and mark-up tags and removing information about the reviews that are not required for sentiment analysis, such as review dates and reviewers' names[19].

Review Analysis: The review analysis step analyses the linguistic features of reviews so that interesting information, including opinions and product features, can be identified. This step often applies various computational linguistics tasks to reviews first, and then extracts opinions and product features from the processed reviews.

Sentiment Classification: There are two main techniques for sentiment classification. The symbolic technique uses manually crafted rules and lexicons, the machine learning approach uses unsupervised or supervised learning to construct a model from a large training corpus[11].

Result: The result is how much people's sentiments, evaluations, attitudes, and emotions in percentages.

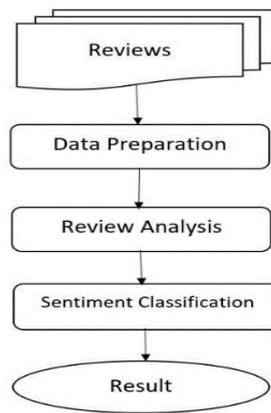


Figure 4: Sentiment Analysis Model

2. LITERATURE SURVEY:

Sentiment analysis or Opinion mining, as it is sometimes called, is one of many areas of computational studies that deal with opinion oriented natural language processing. Such opinion-oriented studies include among others, genre distinctions, emotion and mood recognition, ranking, relevance computations, perspectives in text, text source identification and opinion-oriented summarization [16]. Sentiment analysis has turned out as an exciting new trend in social media with a gamut of practical applications that range from applications in business (marketing intelligence; product and service bench marking and improvement), applications as sub component technology.

[1] Mondher Bouazizi , Tomoaki Ohtsuki; In this case study, the purpose of the multi-class classification of online posts of twitter users, and show how completely it is reasonable to go with the classification, and the limitations and, difficulties of this task have been studied. The intellectual methodology of multi-class classification acquires an accuracy of 60% for seven various sentiment classes which, associated with an accuracy of 81% for binary classification, emphasizes the outcome of having multiple classes on the classification execution. None the less, they proposed a novel model to represent the different sentiments and show how this model helps to understand the relation between various sentiments. The model is then used to analyse the challenges that multi-class classification presents and to spotlight possible future developments to multi-class classification accuracy.

[2] Huizhi Liang , Umarani Ganeshbabu, Thomas Thorne; In this project, a Gaussian Process Dynamic Bayesian Network has been proposed to model the dynamics and interactions of topics on social media such as twitter. This project has conducted experiments on a real-world dataset that was crawled from twitter with 9.72 million tweets. -level sentiment classification approaches can be used to further improve the classification accuracy, thus can promote.

[4] M. Bouazizi, T. Ohtsuki; Proposed a novel method about the classification of sentiments into binary and ternary training which might be prominent classes and additionally, classify text accrued from Twitter into 7 subclasses. For this, they created the SENTA tool for the user to select functions and then run type into 7 different classes (love, anger, hate, fun, happiness, sadness neutral). For tweet Emotion type classification, Random Forest classifier on textual Twitter information is used and 70.1% accuracy on ternary classes, and 60.2% on 7 training has been attained.

[5] Shihab Elbagir Saad , Jing Yang; This study was aimed to accomplish a detailed sentiment analysis of tweets based on ordinal regression applying machine learning techniques. The supposed approach consists of primitive pre-processing tweets and using a feature extraction method that creates an applicable feature. Multicultural logistic regression (SoftMax), Support Vector Regression (SVR), Decision Trees (DTs), and Random Forest (RF) algorithms are given for sentiment analysis classification in the proposed framework.[5], For the actual implementation of this system, a Twitter dataset publicly made acquirable by the NLTK corpora resources are used. Experimental findings share that the proposed approach can discover ordinal regression using machine learning methodologies with good accuracy. Moreover, results indicate that Decision Trees obtain the best results outperforming all the other algorithms. Multinomial logistic regression 67.2%, Support Vector Regression 81.95%, Random Forest 83.2%, Decision Tree 91.81%.

[6] Xianghua Fu , Jingying Yang , Jianqiang Li , Min Fang , Huihui Wang; In this study, Lexicon-enhanced Long short-term memory network (LSTM) has been proposed. The model first uses sentiment lexicon as severely information pre-training a word sentiment classifier and then gets the sentiment embeddings of words including the words not in the lexicon. merging the sentiment establishing and its word establishing can make word representation more accurate. Furthermore [6], The results of experiments on English and Chinese datasets show that our models have comparative or better results than the existing models.

[7] Azwa Abdul Aziz , Andrew Starkey; In this study, a method is known as Contextual Analysis (CA) has been proposed. It is a mechanism that constructs a relationship between words and sources that is constructed in a tree structure detected as Hierarchical Knowledge Tree (HKT). yet, Tree Similarity Index (TSI) and Tree Differences Index (TDI), As a result, the prediction model established indicated estimation error within 2.75 to 3.94 and 2.30 for 3.51 for average absolute differences. [7], furthermore, this method also can cluster sentiment

words into positive and negative beyond having any linguistics resources used and at the equal time capturing changes of sentiment words when a new dataset is applied.

[10] Zhengjie Gao , Ao Feng , Xinyu Song , Xi Wu; In this paper, a new method to learn sentiment-specific word embeddings from Arabic tweets and use them in the Arabic Twitter sentiment classification has been proposed. Moreover, a feature ensemble model of surface and deep features. The exterior features are manually extracted features, and the deep features are generic word embeddings and sentiment specific word embeddings. [10], The extensive experiments are performed to test the effectiveness of the surface and deep features ensemble, pooling functions, embeddings size, and cross-dataset models. The models are discovered on three different datasets of Arabic tweets, and they outcompete the prior results on all these datasets with a suggestive accelerate in the F-score. The developed results demonstrate that:

- 1) the highest performing model is the ensemble of surface and deep features.
- 2) the approach acquires state-of-the-art results on multiple benchmarking datasets.

[12] Guixian Xu , Yueting Meng , Xiaoyu Qiu , Ziheng Yu , Xu Wu; In this study, the results of a comprehensive systematic literature review of the methods and techniques employed in cross-domain sentiment analysis have been presented. This paper focused on studies published between 2010 and 2016. From our analysis of those determines, it is clear that there is never an absolute solution., Hence, one of the aims of this review is to create a resource in the form of an overview of the techniques, methods, and approaches that have been used to attempt to solve the problem of cross-domain sentiment analysis to assist researchers in developing new and more accurate techniques in the future

[16] Ibrahim Said Ahmad, Mohd Ridzwan Yaakub, Mohammad Darwich; In this study, a framework has been proposed. This framework is used for discovering sentiments from tweets of Indian Railways. These types of frameworks are known as domain-specific frameworks. It has different classifiers such as SVM, C4.5, and Random Forest are used to leveraging business intelligence. Precision, recall, F-measure, and accuracy are the measures of the evaluation procedure, the result of Sentiment Analysis of Indian Railway tweets revealed that this framework is very useful. It obtained higher accuracy. Therefore, it can be used in other domains as well.

[14] Dr. Krishna Madhuri; Social feedback is the term defined as “The opinions or sentiments of people exchanges over online social networks or any other online platforms. Before describing the research carried out in this paper, here is the essence of the literature review

made. A framework of opinion mining has been introduced for aspect extraction, opinion classification, summary production, and evaluation. According to this report, opinion mining is a subfield of natural language processing, information retrieval, and text mining.

[17] Chaitanya Bhagat, Deepak T. Mane; The techniques of communalizing user data have been surveyed in this article. This review provides multifaceted insight into the evolution of sentiment analysis into the limelight between the explosion of a plethora of data on the internet.[17], Along with these, various methods of evaluating sentiments have been studied, categorized, and compared. Limitations are also got exposed to provide scope for better research in the future.

[18] Munir Ahmad, Shabib Aftab, Muhammad Salman, Noureen Hameed; This paper emphasizes the various methods for classifying natural language text reviews following opinions expressed in the text to analyse whether the extensive behaviour is negative, positive, or neutral. [18], The method of analysis on polarity classification was discussed in experimental work by using well-known classifiers including Naïve byes, Support vector machine, and Logistic regression for predicting the user reviews.

[20] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu and Bing Qin; States that tweets may be labelled using Senti Word-Net together with Naive Bayes for the category into particular training based totally on their suitability with the subject searched. [20], Twitter is a miniaturized scale blogging website online which gives a stage for people to percentage and specific their views approximately topics, happenings, items, and distinctive administrations. Utilization of Senti Word-Net against Naive Bayes can beautify the exactness of association of tweets, by way of giving energy, cynicism, and objectivity rating of phrases present in tweets.

[21] Penqfei Liu, XipenqQiu, Xuanjing Huanq; States approximately evaluations at the film. It's very complicated to move across all of the views because of the wide variety of opinions frequently posted for a movie on various social network websites, by using sentiment evaluation is feasible to divide overall reviews into superb, bad, and neutral opinions. [21], The proposed system includes system learning strategies for the sentiment analysis, emotions of the humans with the assist of the R language; it could be used for predicting the reputes of the strolling movies. In the sentiment analysis, every single word in the sentence is classified into three types including fine, terrible, and neutral, this technique is implemented using textual content pre-processing strategies and the Naïve Bayes approach for sentiment class.

AUTHOR	TITLE	METHODS/TOOLS	RESULT	DATASET
HUIZHILIANG, UMARANI, GANESHBABU, THOMAS THORNE	A Dynamic Bayesian Network Approach for Analysing Topic-Sentiment Evolution	Gaussian process Regression, Markov Carlo gaussian	predict future sentiment 37 hashtags and 9.72 million tweets in 32 days	Twitter
MONDHER BOUAZIZI, TOMOAKI OHTSUKI	Multi-Class Sentiment Analysis on Twitter: Classification Performance and Challenges	Multi-Class Classification 2D,3D,4D,5D,6D,7 D Sentiment classes.	2D 86.0%, 3D 72.5%, 4D 61.2% 5D 61.8%, 6D 60.4%, 7D 60.2%	Twitter
SHIHAB ELBAIR, JING YANG	Twitter Sentiment Analysis based on ordinal Regression	Multinomial logistic regression, Support Vector Regression Random Forest Decision Tree	LR 67.2%, SVM 81.95% RF 83.2%, DT 91.81%	Twitter
YOUNGSUB HAN, KWANGMI KIM KO	Sentiment Analysis on Social Media Using Morphological Sentence Pattern Model	Movie's review convolutional neural network	Accuracy 91.2%	Movie's review Twitter, IMDB
NORA AL-TWAIRESH AND HADEEL AL-NEGHEIMISH	Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets	BERT MODEL CNN	Accuracy 79.99%	Twitter
DR. KRISHNA MADHURI	A Machine Learning based framework for sentiment classification: India Railways Case Study	C4.5, Naive Bayes, SVM and Random Forest	C4.5 89.5% Naive Bayes 89% SVM 91.5% Random Forest 90.5%	Twitter
ASHUTOSH BANSAL, CHUNNI LAL GUPTA, A MURLIDHAR	A Sentimental analysis for YouTube Data using supervised Learning Approach	Support Vector Machine Decision Tree K Nearest Neighbours	YouTube Accuracy: SVM 93% KNN 80% DT 73%	YouTube
BRIAN KEITH, CLAUDIO MENESES	A Pattern-Based Approach for Multi-class Sentiment Analysis in Twitter.	Random Forest	81.32% for two class, 70.1% for three class and 60.2% for seven classes.	Twitter

	Sentiment analysis of movies on twitter.	Naïve Bayes	Accuracy: 79%	Twitter
Dr. MOHAMMAD ATIQUE, NORA AHMED KHAN	Hashtag-level sentiment classification	SVM Classifier	Automatically generate the overall sentiment separation for a given hashtag in a time period, which definitely differs from conservative's sentence-level analysis	Twitter
ZHENGJIE GAO, AO FENG, XINYU SONG, AND XI WU	Target-Dependent Sentiment Classification With BERT	BERT, Embedding	Average classification accuracy has been pushed to high 70s or mid 80s in percentage	Laptop Twitter Hotel
GUIXIAN XU, YUETING MENG1, XIAOYU QIU, ZIHENG YU, AND XU WU1	Sentiment Analysis of Comment Texts Based on LSTM	LSTM, RNN, CNN, Naive Bayesian Comparison	High accuracy 92.82%	Hotel Review
LIU LIZHEN SONG WEI, WANG, LU JINGLI	A Novel Feature-based Method for Sentiment Analysis of Chinese Product Reviews	feature-based vector model and a novel weighting algorithm	Comparison HADV is 89.1, which is 3% greater than TF 86.1 weighting algorithm, and 2% greater than TFIDF 87.2 weighting algorithm.	Three review data sets. Books, Hotel, Digital Product

Table 1 Performance of various Sentiment Analysis techniques.

3. SENTIMENT CLASSIFICATION TECHNIQUES:

Sentiment classification technique arranged into Machine Learning and Lexicon-Based Approach. The machine learning approach employments famous machine learning algorithms and usages semantic functions. The Lexicon-Based approach is based on a sentiment lexicon, a set of identified and precompile sentiment phrases. It is split right into the dictionary-primarily based method and corpus-based total method. The text classification method using the machine learning techniques can be roughly classified into Supervised and unsupervised learning methods. The supervised technique makes usage of a huge number of labelled training records. The unsupervised techniques are employed when it is hard to find these labelled training records. The Lexicon-Based approach depends upon getting the opinion lexicon which

is utilized to analyse the text. There are two methods in this approach. Dictionary-Based approach which depends on obtaining opinion root words and then examines the dictionary of their synonyms and antonyms. Corpus-Based approach starts with a roots list of opinion words and then obtaining another opinion word in a huge corpus to support in obtaining opinion words with context-specific orientation.

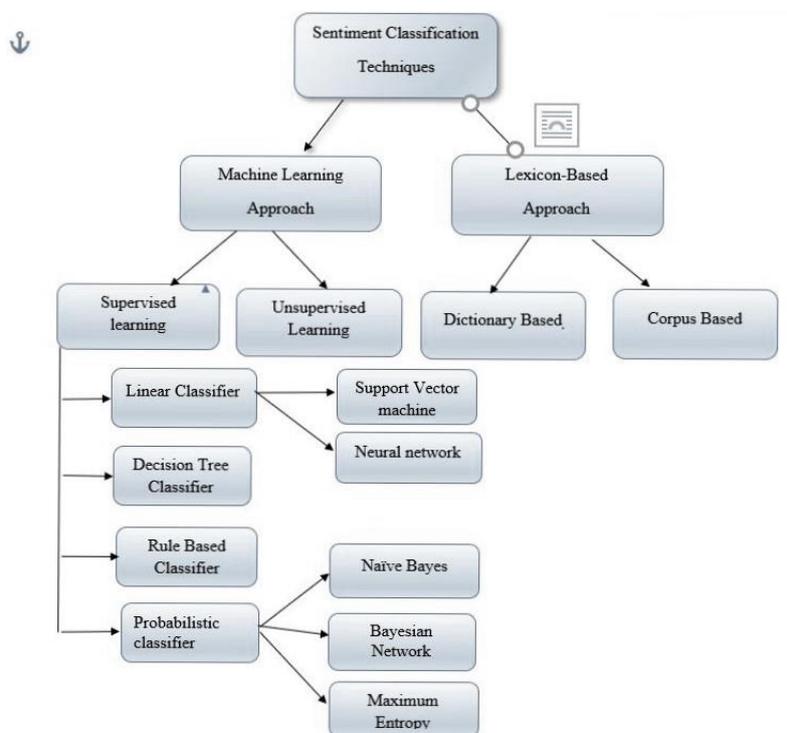


Figure 5: Sentiment Analysis Technique

3.1 Machine Learning Approach: -

This is an automatic classification technique. The classification model is appropriate to the features in the underlying record to categorize according to the class, label. Next for a given instance of an undiscovered class, the model is utilized to predict a class label for it. The tough classification problem is when one label is selected to an instance. The lenient classification problem is when a probabilistic value of labels is selected to an instance. Two types of machine learning supervised and unsupervised learning[18].

3.1.1 Supervised Learning: - The supervised approaches rely upon the occurrence of labelled training data. There are many approaches to supervised some of the most frequently used classifiers in sentiment analysis are Naïve Bayes and Support Vector Machine.

- **Linear Classifier:** - It executes classification based on the linear combination rate of characteristics. Let $W = \{w_1, w_2, w_3, \dots\}$ is word frequency. Vector $C = \{c_1, c_2, c_3, \dots\}$ is linear coefficient vector and S is a scalar then the output of linear predictor will be

$$LP = W.C + S$$

This predictor is called hyperplane which separated two classes[17].

- **Support Vector Machine:** -Support Vector Machine is a discriminative classifier considered because of the suitable sentiment classification approach. It is a statistical

classification technique SVM maps input (actual-valued) feature vectors right into an advanced dimensional feature area via a few nonlinear mappings. SVMs are developed on the precept of structural hazard minimization. The structural hazard minimization seeks to discover a hypothesis (h) for which one can locate the bottom probability of blunders whereas the conventional mastering strategies for pattern reputation are primarily based at the minimization of the empirical chance, that's a try and optimize the performance of the gaining knowledge of set[23]. Computing the hyper aircraft to split the statistics point i.e., Education and Support Vector Machine results in quadratic trouble. Support Vector Machine can study a bigger set of patterns and capable of scale higher, because of class complexity it does not depend on the dimensionality of the function space. Support Vector Machine has the capacity to replace the schooling styles dynamically every time there may be a brand-new sample throughout the category.

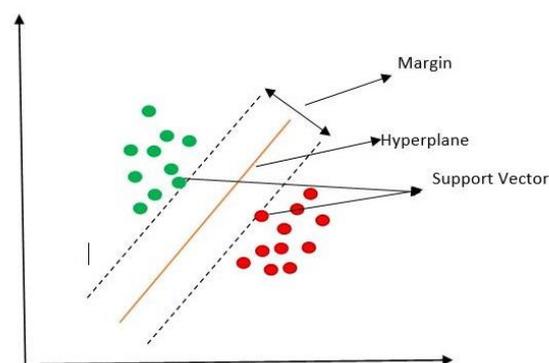


Figure 6: Support Vector Machine

- **Neural Network:** Neural Networks are a class of fashions within the widespread gadget gaining knowledge of literature. Neural networks are a particular set of algorithms that have transformed machine learning. They are inspired by the aid of organic neural networks and the popular so-called deep neural networks have shown to work quite properly. Neural Networks are themselves standard feature processes, that is why they can be hired to almost any device learning trouble regarding studying a complex mapping from the entrance to the output area[16]. Neural Network-based strategies have executed brilliant improvement in an expansion of herbal language processing obligations.
- **Decision Tree Classifier:** -The decision tree algorithm builds the classification model in the form of a tree structure. In this classification, a state is used to divide the data. Data which satisfy the condition is placed in one class and rest data in other class[17]. It is a recursive technique. The decision tree is a restriction less method that creates a model that predicts by simple decision rules which are inferred from the data features. If the tree is deep, the decision rules become more complex.
- **Rule-Based Classifier:** - This classifier builds usage of certain restrain as IF, THEN. It can be written as: - IF condition THEN decision
The restrains can be produced throughout the training phase calculate on our requirements.

- **Probabilistic Classifier:** - This classifier is capable to predict a probability function over a set of classes for a given input data. It does not give only the maximum probable classes but a probability function over all classes.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability
Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Figure 7: Probabilistic Classifier

- **Bayesian Network:** - This model expects that there is a capable dependency between appearance. It is a managed acyclic graph in which nodes represent irregular variables and edges represent dependencies. It is a very luxurious model, therefore is not used frequently.
- **Maximum Entropy:** - The maximum entropy is a classifier that calculates the exponential models. Unlike the Naïve Bayes classifier, this classifier does not calculate on the expectation that the features are conditionally separate from each other.

3.1.2 Unsupervised Learning: It does no longer consist of a category and they do not provide the correct objectives in any respect and as a result depends on clustering. unsupervised learning approach no training dataset provided.

3.2 LEXICON-BASED APPROACH: -

Opinion phrases are used in lots of sentiment class assignments. Positive opinion words are applied to reveal some favoured activities, even as terrible opinion phrases are hired to specify a few undesired events. There are also opinion phrases and idioms which organized are known as opinion lexicon. There are 3 crucial methods on the way to deliver collectively or accumulate the opinion phrase list. The manual method may be very time arduous and it isn't used alone. It is typically joined with the alternative computerized procedure as a very last check to keep away from the errors that resulted from automated techniques[23].

- **Dictionary-based approach:** - A small set of opinion words is obtained manually with known orientation. Then, this set is increased with the aid of searching inside the well-known corpora WordNet or word list for their synonyms and antonyms. The newly obtained words are mixed to the basic list then the subsequent iteration starts off evolved. The iterative system stops whilst no new words are deleted. After the process is finished, the manual examination can be taken out to eliminate or correct errors. The dictionary-primarily based approach has the principal downside that is the weak spot to locate opinion words with discipline and context-specific orientations[21].

- **Corpus-Based Approach:** - The Corpus-based method serves to resolve the hassle of acquiring opinion words with context-specific orientations. Its method depends upon syntactic patterns or styles that arise typically along with a root listing of opinion words to determine other opinion words in a big corpus.

4. SENTIMENT ANALYSIS APPLICATIONS AND CHALLENGES:

4.1 Application:

Opinions are central to almost all human activities because they are key influencers of our behaviour. Whenever we need to make a decision, we want to know others' opinions. In the real world, businesses and organizations always want to find consumer or public opinions about their products and services. Individual consumers also want to know the opinions of existing users of a product before purchasing it, and others' opinions about political candidates before making a voting decision in a political election. In the past, when an individual needed an opinion, he/she asked friends and family. When an organization or a business needed public or consumer opinions, it conducted surveys, opinion polls, and focus groups. Acquiring public and consumer opinions have long been a huge business itself for marketing, public relations, and political campaign companies.

4.2 Challenges:

Incremental Approach:

Analysis of real-time data is not one-time operation. Whenever data is further, we essential to do analysis then why should not we use the preceding analysis result. An Incremental approach agrees an existing result to be updated using one new individual data instance, without having to re-process past instances. This may be useful in situations where the entire dataset is not available when the data changes over time.

Parallel Computing for Massive Data

If we split the computation into tasks or processes that can be performed simultaneously, then there can be an improvement in the speed over the use of parallelism, it is necessary to accomplish this in sentiment analysis for immense data of social media, where immense instant messages are published every day so that we can utilize the overall computing power[5].

Credibility/Behaviour

Behaviours in social media are only observed by the traces they leave in social media. We rarely observe the driving factors that cause these behaviours; nor can we interview individuals regarding their behaviour's. Even if a behaviour is analysed on social media and related patterns are gleaned, it's difficult to verify the validity of these behavioural patterns. Evaluation becomes even more challenging for industries in which important decisions are to be made based on observations of individual behaviour.

Sarcasm

Sarcasm can be used to upset or offend or can be used for comic effect. It means false positives for e.g., "Children really brighten up a household - they never turn the lights off". Detecting

sarcasm from the expressions and finding out the correct context-related sentiments is a challenging task. It is an ironic or satirical remark that seems to be praising someone or something but is really taunting or cutting.

Grammatically Incorrect Words

There are many approaches that analyse sentiments but hardly any work accomplished on grammatical errors. The results of sentiment analysis can be enhanced if these types of errors can be mapped to accurate words.

Review Author Segmentation

Opinion towards a target may be specified by many people who can be called as review authors. Depending on the commenting style of these authors, they should be categorized so that integrity evaluation will be easy. In conclusion making this credibility evaluation is helpful.

5. CONCLUSION & FUTURE SCOPE:

This paper presenting a survey of sentiment analysis and classification algorithms. This survey accomplishes that sentiment classification is still an open field for exploration. There is a lot of opportunity for algorithms in it. SVM and Naïve Bayes are the most popular algorithms for sentiment classification. We can see dataset frequently consist of user opinion on social media i.e., Twitter dataset, Facebook dataset or user reviews on the movie which is written on any website. Sentiment analysis of tweets is very popular. Sentiment can be classified into main categories (Positive, Negative, and neutral). Random Forest, Support Vector Machine and Naïve Bayes are machine learning algorithms that give the highest sentiment classification accuracy. The dataset from sites like Amazon, IMDB, Flipkart is generally used in sentiment analysis. The main goal of this paper is to give an overview of the latest updates in sentiment analysis and classification methods and it includes a brief discussion on the challenges of sentiment analysis for which the work needs to be done.

REFERENCES

1. Mondher Bouazizi , Tomoaki Ohtsuki; “Multi-class sentiment analysis on twitter: Classification performance and challenges Big Data Mining and Analytics” IEEE Access PP(99):1-1 , Year- 2019.
2. Huizhi Liang , Umarani Ganeshbabu, Thomas Thorne; “ A Dynamic Bayesian Network Approach for Analysing Topic-Sentiment Evolution” IEEE Access PP 23(6):1-1, Year- 2020.
3. Jianfei Yu , Jing Jiang , Rui Xia ; “Entity-Sensitive Attention and Fusion Network for Entity-Level Multimodal Sentiment Classification” IEEE Xplore pp:1-4, Year-2020

4. M. Bouazizi, T. Ohtsuki; “A Pattern-Based Approach for Multi-class Sentiment Analysis in Twitter” in Proc. IEEE ACCESS, pp. 20617-20639, Year- 2017.
5. Shihab Elbagir Saad , Jing Yang; “Twitter Sentiment Analysis Based on Ordinal Regression”, IEEE Access DOI 10.1109/.2952127, Year -2019.
6. Xianghua Fu , Jingying Yang , Jianqiang Li , Min Fang , Huihui Wang; “ Lexicon-Enhanced LSTM With Attention for General Sentiment Analysis " IEEE ACCESS 10.1109/ 2878425 , Year-2018
7. Azwa Abdul Aziz , Andrew Starkey; “Predicting Supervise Machine Learning Performances for Sentiment Analysis Using Contextual-Based Approaches” IEEE Access PP(99):1-1, Year-2020.
8. Lakshmish Kaushik , Abhijeet Sangwan ,John H. L. Hansen; “ Automatic Sentiment Detection in Naturalistic Audio “ IEEE/ACM Transactions on Audio, Speech, and Language Processing PP25(8):1-1, Year-2017.
9. Nora Al-Twairesh , Hadeel Al-Negheimish; “ Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets” IEEE Access PP(99):1-1, Year: 2019 .
10. Zhengjie Gao , Ao Feng , Xinyu Song , Xi Wu; “Target-Dependent Sentiment Classification With BERT” IEEE Access PP7:1-1, Year-2019.
11. Tareq Al-Moslmi , Nazlia Omar , Salwani Abdullah , Mohammed Albared “Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review” IEEE Access PP16173 – 16192, Year- 2017.
12. Guixian Xu , Yueting Meng , Xiaoyu Qiu , Ziheng Yu , Xu Wu; “Sentiment Analysis of Comment Texts Based on BiLSTM “ IEEE Access PP 51522 – 51532, Year- 2019.
13. Bo Zhang , Duo Xu , Huan Zhang , Meizi Li ; “STCS Lexicon: Spectral-Clustering-Based Topic-Specific Chinese Sentiment Lexicon Construction for Social Networks” IEEE Transactions on Computational Social Systems PP 1180 - 1189 , Year- 2019.
14. Dr. Krishna Madhuri ; “A Machine Learning based Framework for Sentiment Classification: Indian Railways Case Study” ISSN: 2278-3075, Volume-8 Issue-4, Year- 2019.

15. Ankur Goel, Jyoti Gautam, Sitesh Kumar; “Real Time Sentiment Analysis of Tweets Using Naive Bayes” in 2nd International Conference on Next Generation Computing Technologies, Dehradun, Year-2016.
16. Ibrahim Said Ahmad, Mohd Ridzwan Yaakub, Mohammad Darwich; “Beyond Sentiment Classification: A Novel Approach for Utilizing Social Media Data in Business Intelligence” International Journal of Advanced Computer Science and Applications 11(3), Year-2020.
17. Chaitanya Bhagat, Deepak T. Mane; “Survey On Text Categorization Using Sentiment Analysis” International Journal of Scientific & Technology Research 8(08):1189-1195, Year-2019.
18. Munir Ahmad, Shabib Aftab, Muhammad Salman, Noureen Hameed; “Sentiment Analysis using SVM: A Systematic Literature Review” International Journal of Advanced Computer Science and Applications 9(2), Year- 2018.

19. M. Indhraom Prabha, G. Umarani Srikanth; “Survey of Sentiment Analysis Using Deep Learning Techniques” 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Year- 2019.
20. Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu and Bing Qin; “Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification”, Association for Computational Linguistics PP 1555–1565, Year-2014.
21. Penqfei Liu, XipenqQiu, Xuanjing Huanq; “Research on Short Text Classification Algorithm Based on Neural Network” 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, pp. 1726-1729, Year-2018.
22. Paolo Ferragina, Francesco Piccinno, Roberto Santoro, —On Analyzing Hashtags in Twitter, in Dipartimento di Informatica University of Pisa Proceedings of the Ninth International AAI Conference on Web and Social Media, 2016.
23. Duyu Tang, Bing Qin, Ting Liu; “ Document Modeling with Gated Recurrent Neural Network for Sentiment Classification” Conference on Empirical Methods in Natural Language Processing p1422–1432 Year-2015.
24. B. Plank and D. Hovy; “Approaches to Subjectivity, Sentiment and Social Media Analysis” in Proc. of the 6th Workshop on Computational, Year-2015.