

# Machine Learning Techniques for study Sentiment Analysis of Human Thoughts

<sup>1</sup> Kuna Durga Pavan Desa Rao, <sup>2</sup> M. Rama Bhadra Rao

<sup>1</sup> MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram,  
Narsapur, Andhra Pradesh 534280

[pavankuna446@gmail.com](mailto:pavankuna446@gmail.com)

<sup>2</sup> Assistant Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram,  
Narsapur, Andhra Pradesh 534280

**Abstract:** *The objective of our studies is to analyze how gadget getting to know strategies can be applied to investigate the emotions and attitudes expressed in human wondering. The take a look at amassed statistics from social media structures and on-line forums, which included a mixture of high-quality, bad, and impartial sentiments expressed with the aid of customers. Several device gaining knowledge of algorithms, together with Naive Bayes, SVM, RNNs, CNNs and LSTM Networks, have been employed for sentiment evaluation of the information. The look at observed that the performance of those algorithms various depending at the form of facts being analyzed, with some algorithms performing higher for brief texts such as tweets, even as others worked better for longer texts inclusive of news articles. Additionally, the examiner found that combining more than one algorithm may want to enhance the accuracy of sentiment evaluation. According to the findings, it seems that the usage of machine gaining knowledge of techniques can serve as an amazing manner of scrutinizing human mind and feelings that may have implications for a variety of applications, which includes advertising, politics, and intellectual fitness. This article presents a comprehensive and prepared review of sentiment evaluation strategies. The motive of the evaluation is to investigate and categorize to be had techniques at the same time as evaluating their strengths and weaknesses. The aim is to gain a deeper understanding of the demanding situations that exist within the field and to become aware of capacity solutions and future guidelines. To facilitate this evaluation, we additionally introduce numerous factors that may be used to evaluate the blessings and downsides of every technique inside its category.*

**Keywords**— *Sentiment Analysis, Machine Learning, Classification, Thoughts, Decision Making*

## I. INTRODUCTION

Sentiment analysis or opinion mining is a

technique used to extract and analyze human being's feelings and attitudes in the direction of a particular topic or product

from text information. With the explosion of social media, sentiment evaluation has become increasingly important as agencies and people are searching for to understand the human being's opinion about their merchandise, offerings, or thoughts.

In this article, we can discover how gadget getting to know strategies can be used to carry out sentiment evaluation on human mind. We will delve into various techniques and algorithms used in sentiment analysis, including traditional strategies consisting of rule-based totally techniques and contemporary methods together with deep mastering.

By the give up of this article, readers will have complete information of the ultra-modern strategies used in sentiment analysis and how they can be applied to research human thoughts accurately.

Sentiment analysis is a methodical process of detecting, extracting, measuring, and analyzing emotions and subjective content material through the usage of natural language processing, computational linguistics, textual content analysis, and biometric evaluation. This technique is normally employed to assess "voice of the customer" substances, which includes critiques and survey responses, in addition to healthcare substances and online and social media records. The emergence of

deep language fashions, like RoBERTa, enables sentiment evaluation in greater hard statistics domain names, including news texts, where authors may not explicitly explicit their critiques. Due to the good sized extent of user-generated textual information on-line, sentiment evaluation has turn out to be a hard issue with the fast rise of social media. This studies delves deeply into machine getting to know algorithms and techniques for sentiment evaluation, that are greater adaptive to converting inputs. These algorithms use unigrams, bigrams, and n-grams for facts labelling and processing. Machine gaining knowledge of strategies are usually used for binary category and predicting sentiments as superb or negative, now and again such as other kinds of sentiment as properly, as illustrated within the determine below (Fig 1).

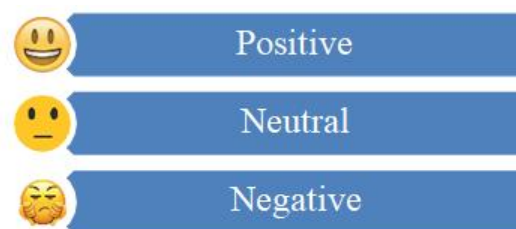


Fig. 1. In General Sentiments Classifications

The three primary categories of sentiment categorization strategies are gadget gaining knowledge of, dictionary-based totally techniques, and hybrid strategies. Well-recognised ML strategies are utilized in system studying strategies, together with

language functions. Mood dictionaries are used inside the dictionary-based totally approach. (Collections of recognised, precompiled mood terms). They may be damaged down into corpus-based approaches and dictionary- primarily based strategies that discover sentiment polarity the use of statistical or semantic techniques. The hybrid approach, which includes both strategies, is popular and most strategies heavily depend upon mood lexicons. Fig2 suggests the recognition of sentiment evaluation in studies area in last 13 years.

## II DETAILED SURVEY OF ML TECHNIQUES

To perform effective comparison among machine learning algorithms used for thoughts classification, we performed detailed survey. In this we selected 34 qualitative research papers published in SCI journals from 2010 to 2021. Findings and classification outcome by survey is shown in following table 1.

TABLE I DETAILED SURVEY OF ML TECHNIQUES FOR THOUGHTS CLASSIFICATIONS

S. No.	SURVEY OUTCOME		
1.	Reference	SCIJournal	Algorithms Used
	Hassan A and Radev D (2010)	Computational Linguistics	Markov Random Walk Model
	<p><b>Findings:</b> Using a Markov random walk model on a detailed graph of word connections, you can create a measurement of polarity for individual words. This model has a significant advantage in that it can quickly and accurately determine a word's polarity, including its direction and intensity. This approach can be used in both supervised scenarios, where a set of labeled words is available for training, and unsupervised scenarios, where only a few seed words define the two polarity categories. The effectiveness of the model is assessed through experiments using a collection of positively and negatively labeled words.</p> <p><b>Classification Outcomes:</b> Positive and Negative Classifications Of Sentiments</p>		
2.	Reference	SCIJournal	Algorithms Used
	Kisioglu P and Topcu YI (2011)	Expert System Applications	Bayesian Belief Network
	<p><b>Findings:</b> The aim of this research was to use a Bayesian Belief Network to detect which customers are likely to leave a telecommunications company. To achieve this, data from a Turkish telecommunication provider was collected. Since the Bayesian Belief Network only works with discrete variables, continuous variables were transformed into discrete variables using the CHAID (Chi-squared Automatic Interaction Detector) algorithm. Additionally, a causal map was created as the foundation of the Bayesian Belief Network, based on the results of correlation analysis, Multicollinearity tests, and expert opinions.</p>		
3.	Reference	SCIJournal	Algorithms Used
	Chen LS et al. (2011)	Journal of Infometrics	A method that uses neural networks to merge the benefits of machine learning methods and semantic orientation index (SOI).
	<p><b>Findings:</b> The effectiveness of semantic orientation indexes is limited, however, they are able to produce results rapidly. On the other hand, machine learning approaches offer more accurate classification, but necessitate a significant amount of training time. To harness the benefits of both methods, a neural network centered method was suggested in this research.</p> <p><b>Classification Outcomes:</b> Positive and negative classes for blogs</p>		
4.	Reference	SCIJournal	Algorithms Used
	Wan X (2011)	Computational Linguistics	Co-training approach outperformed over basic methods(including lexicon-based methods and corpus-based methods) and transductive methods.

7.	<b>Reference</b>	<b>SCIJournal</b>	<b>Algorithms Used</b>
	Ren F, Kang X (2013)	Computer Speech and Language	Hierarchical Bayesian Network, SVM, Naive Bayes
<p><b>Findings:</b> In this study, the researchers utilized Hierarchical Bayesian networks to generate the hidden topic and emotion variables. The suggested approach, which aims to detect a solitary emotion, demonstrates superior performance compared to conventional supervised machine learning models such as SVM and Naive Bayes. Additionally, the authors were able to obtain promising outcomes in detecting intricate emotions using another model. They put the model to the test on a dataset called Ren-CECps, comprising 1487 Chinese blog articles.</p> <p><b>Classification Outcomes:</b> Emotions: Joy, hate, love, sorrow, anxiety, surprise, anger, Expect</p>			
8.	<b>Reference</b>	<b>SCIJournal</b>	<b>Algorithms Used</b>
	Moraes R, Valiati JF, Neto WPG (2013)	Expert System Applications	SVM and ANN
<p><b>Findings:</b> The authors conducted an empirical study to compare the effectiveness of SVM and ANN in document-level sentiment analysis. They examined the necessary conditions, the models generated, and the circumstances under which each approach delivered higher levels of accuracy in classification. They used a standard evaluation framework, including common supervised methods for feature selection and weighting in a conventional bag-of-words model. SVM was found to outperform ANN.</p> <p><b>Classification Outcomes:</b> Positive and negative documents</p>			

Comparative

Analysis was performed on the parameters of advantages, drawbacks and assessment analysis among machine learning algorithms used in above survey of table 1. Following figure (Fig 3) shows the machine learning techniques extracted from this study.

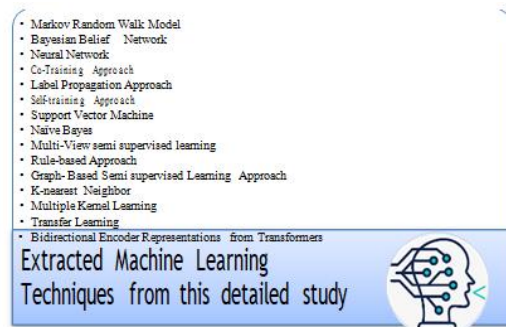


Fig. 3. Extracted Machine learning techniques from this study

Following table summarizes the comparative analysis of ML Techniques from this study.

TABLE II COMPARATIVE ANALYSIS OF ML TECHNIQUES FOR THOUGHTS CLASSIFICATIONS

<p><b>Classification Outcomes:</b> Opinion Classifications positive, negative and neutral.</p>			
15.	<b>Reference</b>	<b>SCIJournal</b>	<b>Algorithms Used</b>
	Hajmohammadi MS, Ibrahim R, Selamat A (2015)	Computer Science	Graph-based semi-supervised learning approach
<p>Proposed a fresh approach that combines sentiment information from labeled and unlabeled data through a semi-supervised learning technique using graphs. The objective is to integrate the underlying structure of the unlabeled data in the target language during the learning process. The effectiveness of this model was evaluated using book review datasets in two different languages, and the experimental results revealed that it improves the performance of cross-lingual sentiment classification compared to several other benchmark methods.</p> <p><b>Classification Outcomes:</b> Positive Documents and Negative Documents</p>			
16.	<b>Reference</b>	<b>SCIJournal</b>	<b>Algorithms Used</b>
	Alfaro C, Cano-Montero J, Gómez J, Moguerza JM, Ortega F (2016)	Annals of Operations Research	k-nearest neighbor, support vector machine
<p><b>Findings:</b> To perform sentiment analysis and opinion mining, it is suggested to merge both supervised ML algorithms and unsupervised ML techniques.</p> <p><b>Classification Outcomes:</b> Messages having positive, neutral and negative opinion</p>			

III COMPARATIVE ANALYSIS OF ML TECHNIQUES

S. No.	Machine Learning Techniques Learning Algorithm	Advantages	Drawbacks	Assessment Analysis
1.	Markov Random Walk Model (Ref 1)	1. The model is highly versatile and produces sequences that resemble real-world usage, provided that it accurately reflects operational behavior. 2. The model is founded on a structured stochastic process, for which there exists an analytical theory.	As additional states and interactions between states are introduced, the situation becomes increasingly intricate.	This can be employed to examine various decision scenarios, including marketing uses that center on customers' loyalty to a specific brand of product, shop, or provider.
2.	Bayesian Belief Network (Ref 2,7)	1. Need only a small amount of instruction to begin working. 2. Take minimal time and effort when building the model.	Capable of handling few continuous variables	Even with small training data, achieve good accuracy
3.	Neural Network (Ref 3, 8,9,18,19, 24,28,30, 31,33)	1. Good performance against noise in data 2. Quick execution time	1. Difficult implementation and Interpretation 2. High	1. It takes longer to train than others 2. Technique

4.	Co-Training Approach (Ref 4,21)	Achieve high classification accuracy with a very limited number of labeled data	1. Poor performance on datasets with only one unique view 2. Many features must be available for optimal performance.	1. Very sensitive to data 2. Different Accuracy for Simple and Complex Domains
5.	Label Propagation Approach with twitter follower graph (Ref 5)	Advantages in terms of how quickly it runs and how little knowledge of the structure is needed in advance (no parameters are needed).	The disadvantage of this is that it only generates an amalgam of different answers.	A semi-supervised learning algorithm that works well.
6.	Self-training Approach (Ref 6)	1. Ease of the technique 2. There is no dependence on a classification model.	1. There is a chance to strengthen the input sample if it contains an error. 2. Alert to anomalies	Traditional self-training techniques function poorly.
7.	Support Vector Machine (Ref 7,8,14, 16,19,21, 25,28,30)	1. Training that is relatively simple 2. The ability to generalize well in both theory and practice 3. Not being highly reliant on the number of features in a dataset.	1. You must select the proper Kernel function. 2. A slowdown caused by a rise in the sample size 3. Interpretation issue	1. Excellent results from the experiment. 2. Outperforming the alternatives in terms of benefits

#### IV ASSESSMENT OF STUDY

As according to the special survey, there are numerous algorithms that may be used for sentiment analysis and the fine one relies upon on the unique use case and facts set. Here are some assessed algorithms by using the researchers for sentiment analysis:

1. Naive Bayes: This algorithm is easy and green, making it a famous choice for sentiment evaluation. It calculates the possibility of a file belonging to a specific sentiment category based totally at the frequency of words in the document.
2. Support Vector Machines (SVMs): This

algorithm works by way of finding a hyper plane that separates the wonderful and terrible sentiment facts factors. SVMs are especially powerful whilst coping with high-dimensional information sets.

3. Recurrent Neural Networks (RNNs): RNNs are a form of deep getting to know set of rules that can analyze textual content facts by way of processing it sequentially. This makes them in particular beneficial for studying sequences of textual content, including film opinions or social media posts.

4. Convolution Neural Networks (CNNs): CNNs are another kind of deep studying set of rules that may be used for sentiment analysis. They paintings with the aid of figuring out essential features in the textual content, along with words or phrases, and the usage of them to classify the sentiment.

5. Long Short-Term Memory Networks (LSTMs): LSTMs are a sort of RNN that may better take care of long sequences of textual content statistics, making them beneficial for sentiment evaluation of longer texts, such as purchaser critiques or information articles.

It's essential to word that the overall performance of those algorithms can range depending on the facts set and the unique hassle being solved. It's often an amazing idea to test with distinctive algorithms to look which one works satisfactory for your

particular use case.

## V CONCLUSION

This paper affords an intensive investigation into the categorization of thoughts using machine studying techniques and accomplished evaluation amongst them. Summary of labour is represented amongst table 1 and table 2. From this Survey, the parameters worried/utilized by Researchers are:

A. Mostly used techniques are Support Vector Machine (SVM), Neural Network (NN), Naïve Bayesian algorithms, k-nearest Neighbour (KNN), Long Short-Term Memory (LSTM), Bidirectional Encoder Representation from Transformers (BERT) , Hybrid Algorithms and okay-way clustering and so on.

B. Mostly used dataset are taken from Twitter, Face book, Senwave, Big Five, MBTI, IMDB, Amazon, on line repositories, and actual field facts from particular region.

C. Mostly Used parameters are Accuracy, F1 Score, Recall, RMSE and many others.

The principal studies hole what we diagnosed is that existing research are in the main focusing the sentiment of someone towards outdoor (approximately product, about subject matter etc.). In destiny research work I am going to

introduce the implementation of sentiment evaluation for a person itself to perceive his balance of thoughts to take choices. For this purpose, it is required to increase standard outcomes of Sentiment analysis. To gain the objective of classification of thoughts in this element there may be need to expand a novel classified dataset in an effort to be similarly examined, skilled and tested the usage of gadget mastering strategies.

## REFERENCES

1. Hassan A, Radev D (2010) Identifying text polarity using random walks. In: Proceedings of the 48th annual meeting of the association for computational linguistics. Association for Computational Linguistics, pp 395–403
2. Kisioglu P, Topcu YI (2011) Applying Bayesian belief network approach to customer churn analysis: a case study on the telecom industry of Turkey. *Expert Syst Appl* 38(6):7151–7157
3. Chen LS, Liu CH, Chiu HJ (2011) A neural network based approach for sentiment classification in the blogosphere. *Journal of informatics* 5(2):313–322
4. Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.
5. Spurious M, Sudan N, Upadhyay S, Baldrige J (2011) Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the first workshop on unsupervised learning in NLP. Association for Computational Linguistics, pp 53–63
6. He Y, Zhou D (2011) Self-training from labelled features for sentiment analysis. *Information Processing and Management* 47(4):606–616
7. Ren F, Kang X (2013) Employing hierarchical Bayesian networks in simple and complex emotion topic analysis. *Computer Speech and Language* 27(4):943–968
8. Moraes R, Valiati JF, Neto WPG (2013) Document-level sentiment classification: an empirical comparison between SVM and ANN. *Expert System Applications* 40(2):621–633
9. Kalchbrenner N, Grefenstette E, Blunsom P (2014) a convolution neural network for modelling sentences. In: Proceedings of the 52nd annual meeting of the association for computational linguistics
10. Hajmohammadi MS, Ibrahim R, Selamat A (2014) Cross-lingual sentiment classification using multiple source languages in multi-view semi-supervised learning. *Eng Appl Artif Intell* 36:195–203
11. Li G, Liu F (2014) Sentiment analysis based on clustering: a framework in improving accuracy and recognizing

neutral opinions. Appl Intell 40(3):441–452.

12. Prasadu Peddi (2015) "A review of the academic achievement of students utilising large-scale data analysis", ISSN: 2057-5688, Vol 7, Issue 1, pp: 28-35.

13. Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.