Vol.15, Issue No 2, 2025

Sentiment Insights a Study of Humans Emotions using Machine Learning Techniques

[1]MULAPARTHI SRILAKSHMI, [2] KOLLATI VISHNU VARDHINI,[3] THONTA DURGA RAO, [4] KALIGITHI VIVEK, [5]Mrs. V. NAVYA DEVI,

[1234]Student Department Of CSE, DNR College Of Engineering & Technology, Balusumudi, Bhimavaram, India.

[5]Assistant Professor, Department Of CSE, DNR College Of Engineering & Technology, Balusumudi, Bhimavaram, India.

Abstract—

We want to find out how to use machine learning techniques to decipher the sentiments and beliefs conveyed in human thought. Positive, negative, and neutral user attitudes were all included of the data set collected from online forums and social media sites for the research. For the purpose of sentiment analysis, a number of machine learning techniques were used, such as RNNs, CNNs, LSTM Networks, Naive Bayes, and Support Vector Machines. Depending on the data type, the algorithms' performance varied; for example, some algorithms did better with shorter texts (like tweets) and others with longer texts (like news articles), according to the research. Combining many algorithms might enhance sentiment analysis accuracy, according to the research. Findings suggest that machine learning approaches may be useful for studying human emotions and ideas; this has broad implications for fields including politics, marketing, and mental health. In this post, we will take a look at sentiment analysis techniques in detail. Examining and classifying existing methods while contrasting their advantages and disadvantages is the goal of the review. The goal is to learn more about the problems in the field so we can figure out how to fix them and where to go from here. In order to make this analysis easier, we include a number of criteria that may be used to weigh the pros and cons of each approach in the category.

Keywords— Sentiment Analysis, Machine Learning, Classification, Thoughts, Decision Making I.

INTRODUCTION

The goal of sentiment analysis, also known as opinion mining, is to glean from text data the feelings and thoughts that individuals have about a certain subject or product. Businesses and individuals alike are increasingly turning to sentiment analysis in an effort to gauge public perception of their goods and services in light of the meteoric rise of social media. Here, we'll take a look at how sentiment analysis on human ideas may be done using machine learning approaches. In this article, we will explore the many algorithms and methodologies used for sentiment analysis, from more conventional approaches like rule-based systems to more cutting-edge ones like deep learning. Readers will walk away from this essay with a firm grasp of the cutting-edge methods for sentiment analysis and how to put them to work for precise human thinking analysis. Location: Lovely Professional University Phagwara, Punjab, India 979-8-Dr. Geeta Sharma School of Computer Applications Email: geeta.26875@lpu.co.in Using tools from the fields of natural language processing, computational linguistics, text analysis, and biometric analysis, sentiment analysis systematically detects, extracts, measures, and examines subjective content and emotions. A number of fields and types of information, including healthcare, internet and social media data, and "voice of the customer" materials like reviews and survey replies, often use this method for evaluation. New deep language models, such as RoBERTa, make sentiment analysis possible even in difficult data domains like news texts, where writers may not be so forthcoming with their views. With the proliferation of social media comes a deluge of user-generated textual data, making sentiment analysis a challenging task. More flexible in response to new or altered inputs, machine learning algorithms and methods for sentiment analysis are the focus of this study. For data labeling and processing, these algorithms use unigrams, bigrams, and n-grams. Binary classification and sentiment prediction using machine learning algorithms is common, as seen in the image below (Fig 1), however additional types of sentiment may also be included.

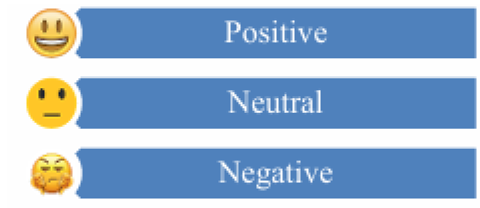

Fig. 1. In General Sentiments Classifications

Generally speaking, there are three distinct approaches to sentiment classification: hybrid, dictionary-based, and machine learning. In machine learning approaches, well-known ML algorithms and language characteristics are used. The dictionary-based approach makes use of mood dictionaries. (sets of pre-existing, recognized mood words). Methods that determine the polarity of a sentiment by statistical or semantic means fall into two broad categories: those that rely on corpora and those that rely on dictionaries. Most methods depend significantly on mood lexicons, and the hybrid approach that combines the two is rather common. Figure 2 displays the research area's adoption of sentiment analysis during the last thirteen years.

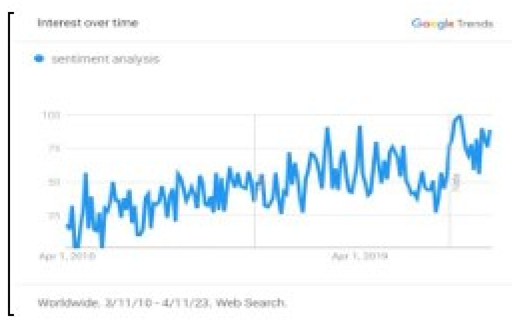

Fig. 2. Google Trends result for 'Sentiment Analysis' of last 13 years

In this study, we conducted a detailed survey in Chapter 2 after outlining the concept of survey in Chapter 1. In Chapter 3, we take a close look at the various machine learning techniques that have been used for opinion mining and sentiment analysis. Chapter 4 presents the study's evaluation, and Chapter 5 wraps up our work and discusses the survey's potential future applications.

DETAILED SURVEY OF ML TECHNIQUES

We conducted an extensive study to compare several machine learning techniques used for thinking categorization. From 2010 to 2021, we culled 34 qualitative research articles from SCI publications for this. The results and categorization based on the survey are shown in table 1.

TABLE I DETAILED SURVEY OF ML TECHNIQUES FOR THOUGHTS CLASSIFICATIONS

S. No.	SURVEY OUTCOME		
1.	**Reference**	**SCI Journal**	**Algorithms Used**
	Hassan A and Radev D (2010)	Computational Linguistics	Markov Random Walk Model
	Findings: 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.		
	Classification Outcomes: Positive and Negative Classifications Of Sentiments		
2.	**Reference**	**SCI Journal**	**Algorithms Used**
	Kisioglu P and Topcu YI (2011)	Expert System Applications	Bayesian Belief Network
	Findings: 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.		

	Classification Outcomes: Customer churn analysis on parameters Place of residence, Age, Tenure, Tariff type, Average billing amount, Trend in billing amount, Averageminutes of usage, Average frequency of usage and a dependent variable churn.		
3.	**Reference**	**SCI Journal**	**Algorithms Used**
	Chen LS et al. (2011)	Journal of Informetrics	A method that uses neural networks to merge the benefits of machine learning methods and semantic orientation index (SOI).
	Findings: 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.		
	Classification Outcomes: Positive and negative classes for blogs		
4.	**Reference**	**SCI Journal**	**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.

			methods.
	Findings: Suggest utilizing a bilingual co-training strategy that incorporates both English and Chinese perspectives by utilizing more unlabeled Chinese data. The effectiveness of the suggested approach was demonstrated through experiments on two test sets, where it performed better than basic and transductive methods.		
	Classification Outcomes: Positive and negative sentiments		
5.	**Reference**	**SCI Journal**	**Algorithms Used**
	Speriosu M and Sudan N et al.(2011)	Computational Linguistics	Label Propagation approach with twitter follower graph
	Findings: The suggested method employs label propagation to integrate labels from a maximum entropy classifier that was trained on imprecise labels, knowledge about the types of words stored in a lexicon, and the Twitter follower graph. The results of tests on different datasets for polarity classification reveal that our label propagation technique is similar in performance to a model trained on marked tweets within the same field, and it surpasses both the imperfect supervised classifier that it utilizes and a polarity ratio classifier that is based on a lexicon.		
	Classification Outcomes: Positive and negative opinions		
6.	**Reference**	**SCI Journal**	**Algorithms Used**
	He Y, Zhou	Information	Self-training

COMPARATIVE ANALYSIS OF ML TECHNIQUES

Comparative analysis were 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.

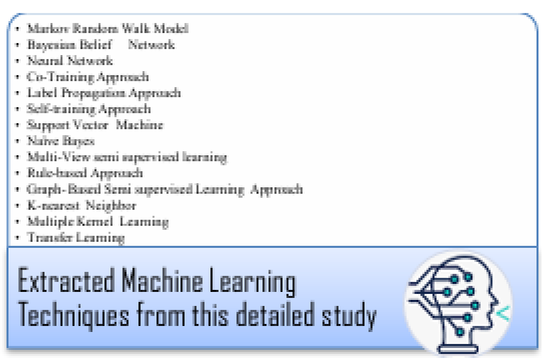

Fig. 3. Extracted Machine learning techniques from this study Following table summarizes the comparative analysis of ML Techniques from this study.

TABLE II COMPARATIVE ANALYSIS OF ML TECHNIQUES FOR THOUGHTS CLASSIFICATIONS

S. No.	Machine Learning Techniques Learning Algorithm	Advantages	Drawbacks	Assessment Analysis
1.	MarkovRandom 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 executiontime	1.Difficult implementation and Interpretation 2. High memory usage	1. It takes longer to train than others Technique 2. Convolutional neural networks are

				a viable alternative to overcome expensive
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

ASSESSMENT OF STUDY

The most effective algorithm for sentiment analysis is going to vary from one use case and data set to another, according to the comprehensive study. The researchers' evaluations of sentiment analysis algorithms are as follows: One prominent method for sentiment analysis is Naive Bayes since it is both simple and efficient. Using the document's word frequency, it determines the likelihood of the document belonging to a certain emotion group. Second, Support Vector Machines (SVMs) distinguish positive and negative sentiment data points using a hyperplane. When working with datasets that have a large number of dimensions, SVMs shine. 3. RNNs: RNNs are a subset of deep learning algorithms that can sequentially process text input for analysis. Because of this, they are great for evaluating long passages of text like social media postings or movie reviews. CNNs, or convolutional neural networks, are a fourth category of deep learning algorithms that include sentiment analysis potential. One way they do this is by looking for key words or phrases in the text and determining which ones best represent the tone. 5. LSTMs: A kind of RNN, LSTMs are better at processing lengthy text sequences, which makes them great for sentiment analysis of lengthier texts like news stories or customer evaluations. The data collection and the particular 783 issue at hand may have a significant impact on how well these algorithms work. Trying out many algorithms is a great way to find the one that suits your needs the most.

Vol.15, Issue No 2, 2025

CONCLUSION AND FUTURE WORK

Extensive research on the classification of ideas using machine learning methods and comparisons between them are presented in this work. Table 1 and table 2 include the work summary. Researchers used the following criteria derived from this survey: A few of the most common methods are Support Vector Machine (SVM), Neural Network (NN), Naïve Bayesian algorithms, K-Nearest Neighbor (KNN), Long Short-Term Memory (LSTM), Bidirectional Encoder Representation from Transformers (BERT), Hybrid Algorithms, and k-means clustering. B. The most popular datasets are culled from several internet sources and real-world data sets, including Senwave, Twitter, Facebook, Big Five, MBTI, IMDB, Amazon, and online repositories. (C) Parameters like as Accuracy, F1 Score, Recall, RMSE, and others are often used. We found a big hole in the literature since most studies have focused on people's feelings toward things outside of themselves (such as products or topics). In my next studies, I want to present the use of sentiment analysis to determine an individual's mental stability in decision-making. To achieve this goal, it is necessary to expand the typical results of sentiment analysis. In this area, the goal of thought categorization necessitates the creation of a new labeled dataset that can be investigated, trained, and evaluated with the use of machine learning techniques.

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 informetrics 5(2):313–322

[4]. Wan X (2011) Bilingual co-training for sentiment classification of Chinese product reviews. Computational Linguistics 37(3):587–616

[5]. Speriosu M, Sudan N, Upadhyay S, Baldridge 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 labeled 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 convolutional neural network for modelling sentences. In: Proceedings of the 52nd annual meeting of the association for computational linguistics

[10]. Hajmohammadi MS, Ibrahim R, SelamatA(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