ISSN No.: 2455-6130
Special Issue: 2, 2022, pp. 28-36

website: www.ijoetr.com

International Journal of Emerging Trends in Research

Sentiment Analysis based on Machine Learning and Deep Learning

Kalyani P. Sable[1*], Dr. S. L. Satarkar[2]

[1, 2]Deptt. of Computer Science & Engineering, Shri Sant Gajanan Maharaj College of Engineering,Shegaon
[32]Deptt. of Computer Science & Engineering,College of Engg & Technology, Akola

Abstract

Various machine learning algorithms for sentiment analysis are discussed in this study. Machine learning classifiers such as Naive Bayes, Decision Tree, Random Forest, Support Vector Machine, KNN, and deep learning classifiers were used to analyze sentiment. We notice some articles in this section that are assisting young investigators in determining the best path for further study. Various social networking sites, E-commerce sites like Amazon, and social media like Facebook, Twitter, and Instagram are popular platforms for users to express their views on many topics. Sentiment analysis employs a machine learning approach and provides a precise assessment of people's feelings without the need for human intervention. The Sentiment analysis divides the text into three categories: positive, negative, and neutral. As a result, any corporation, institution, an examiner can accept the public's view and take action based on it.

Keywords: SocialMedia, Machine Learning, Deep Learning, Sentiment Analysis, Opinion

1. Introduction

Sentiment classification is the way that examines texts for polarity, ranging from positive to negative using the machine learning approach. Machine learning automatically learns human sentiments. Today, social media is an integral aspect of people's lives; they utilize it to share their opinions on many topics such as politics, film ratings, and advertisements. There are numerous social media platforms available, including Twitter, Facebook, Instagram, and more. They use these social media sites to share their opinions on a variety of issues. Therefore, using the training data set, sentiment analysis analyses the text entered by any individual from a particular location. It evaluates the sentiments of that certain text by understanding the sentiment of such a user.

For example, all over the world uses sentiment analysis when they observe that viewers are complaining about faulty products. Instead of being discouraged by unfavorable feedback, Users managed to capitalize on it by airing new products.

A. Phases of Sentiment Analysis

- Documentation: The entire document is subjected to documentation analysis. A document addressing a particular topic has been included at this stage of categorization. Users believe that comparing two text documents is impossible to conduct in a documentation analysis. In the d documentation phase, sentiment analysis is classified using various machine learning algorithms.
- Sentence: In the sentence Phase, sentiment analysis is strongly associated with a subjective categorization. The goal of sentence-level sentiment analysis is to determine if a sentence's polarity is positive, negative, or neutral. The machine learning classifier is used for sentiment analysis in the sentence analysis phase.
- Relative: In the relative phase, find the relation between two entities in the sentences. For example, My Laptop is looking good but the processing is quite slow. In this example, there is an opinion about the laptop that looks good and polarity is positive but the processing speed is quite slow indicating negative polarity. It is also called aspect-level sentiment analysis.
- Phrase: Comment on any product or item is classified into various phrases that are present in the sentence. The phrase level has some benefits and drawbacks, with the benefit being the specific view of the product, item, or entity. However, because of the context polarities issue, the outcome may not be precise.
- Feature-based: Item characteristics are the key features of any product. In feature-based sentiment analysis is described in the documentation as analyzing the various features for the purpose of determining opinions about any product or item. The retrieved features are used to determine whether the sentiment is positive, negative, or neutral.

2. Related Work

Soumya S. et. al. suggested the machine learning model for sentiment analysis of Malayalam tweets. The author proposed the SVM, RF, and NB algorithms for the classification of positive and negative tweets and for preprocessing using TF-IDF methods [1].

Gamal et. al. proposed the machine learning model for sentiment analysis of Arabic tweets based on various Arabic dialects datasets that were utilized which contains more than 15100 tweets or comments. The final result indicated the polarity of the user tweet. For the cross-validation 10-fold cross-validation was used and compared various machine learning classifiers' performance [2].

Satuluri Vanaja et. al. suggested the machine learning techniques for aspect level sentiment analysis of Amazon customer tweets, and classify those tweets in positive, negative, and neutral polarities [3].

Ezpeleta et. al. suggested the model for the classification of spam emails based on the sentiment analysis and analyze the email contents. The Author utilized the hybrid approach to analyze the spam email contents [4].

Cristian R. Machuca et. al. suggested the machine learning model recognize people's feelings, emotions regarding the coronavirus and classify the positive and negative emotions of users. The Twitter dataset was used for sentiment analysis which contains 52000 tweets in the English language and downloads the hashtag of coronavirus [5].

Mujahid, M. et. al. suggested the machine learning model for sentiment analysis of online education tweets during the covid-19 pandemics. For the preprocessing, BOW, and TF-IDF techniques and for performance analysis various machine learning techniques were used [6].

A. Machine Learning approach

The use of well-known machine learning approaches on textual documents is central to the various machine learning approach in sentiment categorization. Machine learning is particularly practical because it is completely automated and manages massive amounts of textual data. The machine learning approach can be classified into supervised, semi-supervised, and, unsupervised for sentiment categorization.

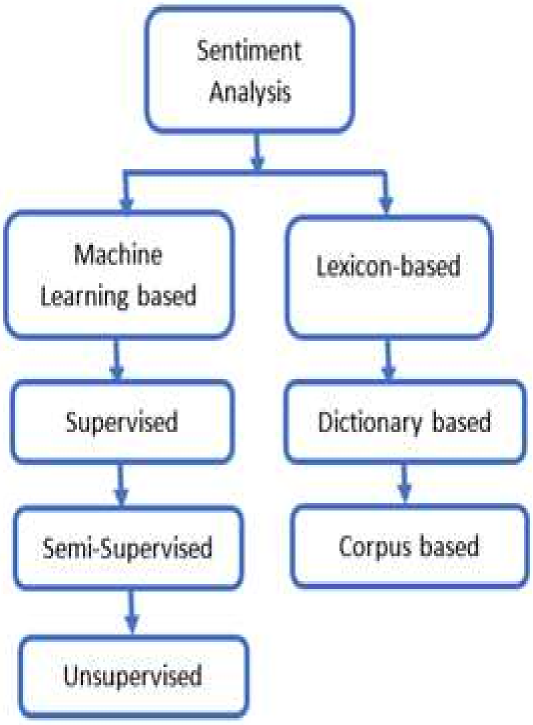

Figure 1: Methods of Sentiment Analysis

B. Steps of Sentiment analysis using Machine Learning

The labeled polarity of any dataset has been used for analyzing the user’s sentiments such as positive, negative, or neutral, etc. The sentiment analysis includes several key processes, including a preprocessing step in which all ambiguous data is deleted. Possible features are obtained from the cleaning database. These characteristics are phrases in the texts that must be transformed into a numeric representation. Text information is converted to a numeric form using vectorization algorithms. A matrix is produced via vectorization, with the column representing a feature set and each row representing a user's review. The classifications method utilizes the matrix as inputs, as well as the cross-validation approach is utilized to select the training and test the data sets. Figure 2 depicts a step-by-step explanation of sentiment analysis.

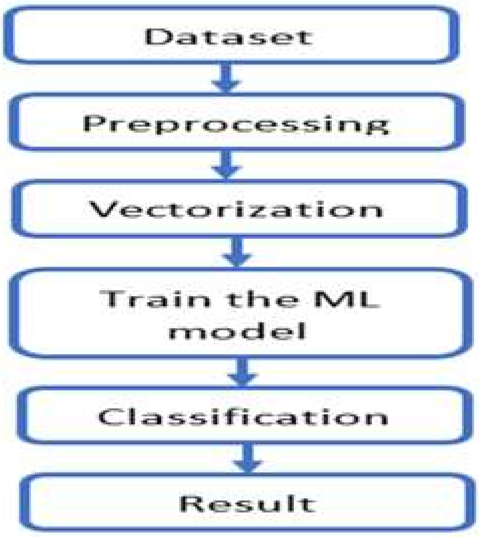

Figure 2: Steps of Sentiment Analysis

Steps for Sentiment analysis

- Dataset: The polarity of the item or product review dataset is used for the sentiment analysis which consists of various positive and negative tweets. The dataset maintained the text data in a .csv file.

- Preprocessing: There is a lot of ambiguous material in users' tweets that must be removed. Initially, the special characters or symbols such as #, %, or @ and unwanted empty spaces are deleted during the preprocessing. Reviewers have been found to frequently repeat the specific character of a word in order to emphasize a statement or to keep their tweets relevant. Stop-words are the majority of phrases in a sentence that do not contribute to any expression. As a result, the next stage in preprocessing entails removing the stop words like is, the, a, an, etc. in a sentence.

- Vectorization: The important features are extracted from the dataset once it has been cleaned in the vectorization step. The features have tokenized the phrases in users' tweets. These phrases must be translated to numeric values in order for user tweets can be represented numerically.

- CountVectorizer: It converts the evaluation into a token counted matrix. To begin, it tokenizes the user's tweets and generates a sparse matrix depending on the number of instances of every token.

- TF-IDF: Its value measures the significance of a phrase in a text file.

- Machine Learning model: The machine learning classifier method can be provided the numerical weight or values as input. For the classification of sentiments, various machine learning methods will be applied.

- Training and Testing Model: For the performance evaluation confusion matrix is generated after the machine learning model has been trained, it shows the total number of positive and negative tweets are accurately estimated, as well as some positive and negative tweets are incorrectly estimated. The predictive performance for each tweet is determined using a confusion matrix, and the ultimate performance is determined by averaging all of the various accuracies.

- Predicted Result: Accuracy, Sensitivity, Specificity, Recall, and F1-Score are the evaluation parameters to measure the performance of the machine learning model. The confusion matrix is created, as well as a table with effectiveness analysis metrics. Lastly, the resulting values are compared with existing techniques.

Table 1: Summary of Related Work for Sentiment Analysis in Machine Learning

Ref. No.	Dataset Used	Language	Tweet Types	Techniques	Accuracy in %
[1]	Malayalam Tweets	Malayalam	Movie tweet	SVM, RF, NB	NB = 92% SVM (linear-kernel) = 94% SVM (RBF-kernel) = 93% RF = 96%
[2]	Arabic dialects	Arabic	Social Media personal tweets	NB, LR, ME, PA, RR, SVM, MNB, Ada-Boost BNB, SGD	PA/RR = 98% SVM/Ada-Boost/LR/BNB = 59%
[3]	E-Commerce	English	Customer review	---	NB = 90% SVM = 83%
[4]	CSDMC-2010 dataset	English	Spam email	Sentiment Analysis	Sentiment analyser = 99% Personality (Sensing) = 99.03% Combination = 99.03%
[5]	Twitter Dataset	English	Covid-19 tweet	Binary Logistic Regression	Binary LR = 78%
[6]	Twitter dataset	English	Covid-19 tweet for online education	SMOTE	SVM = 95% LR = 93% DT = 95% RF = 95% KNN = 62%

C. Deep Learning Approach

The conventional machine learning techniques such as Decision tree, random forest, Support Vector Machine, etc. have performed well in many Natural language processing applications, but have certain weaknesses, which deep learning techniques can solve the issues. Table 2 shows some previous work done in sentiment analysis based on the deep learning approach.

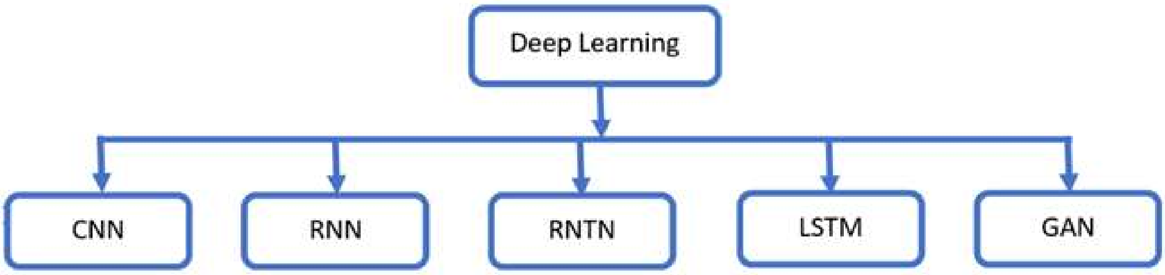

Figure 3: Deep Learning techniques

Table 2: Summary of Related Work for Sentiment Analysis in Deep Learning

Ref. No.	Dataset Used	Language	Tweet Types	Techniques	Accuracy in %
[7]	Twitter Dataset	English	Visual Sentiment Analysis	CNN	CNN = 90%
[8]	Twitter dataset	English	Covid-19 tweet	LinerSVC, BernoulliNB	LinerSVC = 57% BernoulliNB() 47%
[9]	Twitter dataset	English	Covid-19 tweet	Hybrid heterogeneous-SVM, RNN	Hybrid heterogeneous-SVM = 88% RNN = 90%
[10]	Amazon Product tweet dataset	English, French, German, Japanese	Amazon Product tweet	Weakly-Shared DNN	DNN= 80%
[11]	Thai Twitter Data	English	People emotion	DCNN LSTM	DCNN=75% LSTM=75%
[12]	Multi-domain dataset of customer	English	Electronics Product tweet	PNN RBM	PNN= 88% RBM = 90%

3. Challenges in Sentiment Analysis

Parsing: Sentence can be divided into two components, namely subject and object.In sentence which one is verb or adjective and to what it refers seems to be difficult.

Emojis: If the data is in the emojis form, then need to detect whether it is good or bad.

Comparing neutral statement is very much challenging.

Labelled entity recognition: Identification and classification of labelled entities from the text into its pre-defined types is quite challenging.

Rhetorical mode: Analyzed posts containing irony, sarcasm, implication, etc., are tough to identify.

Anaphora Resolution: The problem of resolving what a noun or pronoun refers to.

Tone determination: If the data is in the form of a tone, then it makes really difficult to detect whether the comment is pessimist or optimist.

Social media posts: Finding reviews and opinions from the text which is having lack of capitals, abbreviations, slangs mean informal words, incorrectpunctuation makes sentiment analysis pretty much challenging.

Visual sentiment analysis: In different posts containing textual and visual information, polarities of the sentiment or opinion referred by the given texts may contradict the sentiments of image data.

4. Application

Market Analysis

To determine what is the newest product on the marketplace and what customers desire. After you've done the research, adjust the overall marketing approach accordingly.

Competitive Analysis
To find out what opponents are releasing or what products they have on the marketplace. To research an opponent'sstrategies based on public perception. One of the most common uses of sentiment evaluation is this.

Product Analysis
To find out what customers think regarding your item after it has been released, or analyze the people emotions, reviews in ways you have not seen before. you may quickly assess a customer review by looking for a term for a specific feature.

Social Media Analysis
People share their views on social media in any field like business, government, market, or any other. By the sentiment analysis by searching some keywords you can easily monitor people's sentiment from individual points of view.

Client Feedback
In any commercial industry or corporation, analyzing client feedback is a very important activity. A corporation may readily examine their client's assessment of a product using sentiment analysis, and they can make modifications to their service based on the client's opinion.

3. Conclusion

Collecting the sentiments, views, reviews, tweets, and emotional language and analyzing them, and extracting meaningful information is referred to as sentiment analysis. The need for analyzing and organizing hidden content gathered from various social media platforms like Facebook, Twitter, other social media websites, etc. that contain a massive amount of unstructured data for sentiment analysis. This paper discussed sentiment analysis in the machine and deep learning approach. Several researchers used various machine algorithms such as decision tree, random forest, support vector machine, and Adaboost, Naïve Bayes, Logistic regression, etc., for sentiment analysis. The deep learning approach is made up of several useful and successful methods which are employed to tackle a wide range of issues. In this study, various existing studies are reviewed to provide a thorough understanding of the significant growth of the deep learning area of sentiment analysis.

References

[1] Soumya S. and Pramod K.V., Sentiment analysis of Malayalam tweets using machine learning techniques, ICT Express (2020), DOI:https://doi.org/10.1016/j.icte.2020.04.003.

[2] Gamal, Donia & Alfonse, Marco & El-Horbarty, El-Sayed &M.Salem, Abdel-Badeeh. (2019). Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features. Procedia Computer Science. 154. 332-340. 10.1016/j.procs.2019.06.048.

[3] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 1275-1279, DOI: 10.1109/ICIRCA.2018.8597286.

[4] Ezpeleta, Enaitz; Velez de Mendizabal, Iñaki; Hidalgo, José María Gómez; Zurutuza, Urko (2020). Novel email spam detection method using sentiment analysis and personality recognition. Logic Journal of the IGPL, (), jzz073–. doi:10.1093/jigpal/jzz073.

[5] Cristian R. Machuca et. al. "Twitter Sentiment Analysis on Coronavirus: Machine Learning Approach ", 2021 J. Phys.: Conf. Ser. 1828 012104

[6] Mujahid, M.; Lee, E.; Rustam, F.; Washington, P.B.; Ullah, S.; Reshi, A.A.; Ashraf, I.Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID-19. Appl. Sci. 2021, 11, 8438. https://doi.org/10.3390/app11188438

[7] J. Islam and Y. Zhang, Visual Sentiment Analysis for Social Images Using Transfer Learning Approach, 2016 IEEE Int. Conf. Big Data Cloud Comput. (BDCloud), Soc. Comput. Netw. (SocialCom), Sustain. Comput. Commun., pp. 124130, 2016.

[8] Chakraborty, K., S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. J. A. S. C. Hassanien (2020). "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers-A study to show how popularity is affecting accuracy in social media."97: 106754.

[9] Kaur, H., Ahsaan, S.U., Alankar, B. et al. A Proposed Sentiment Analysis Deep Learning Algorithm for Analyzing COVID-19 Tweets. Inf Syst Front 23, 1417–1429 (2021). https://doi.org/10.1007/s10796-021-10135-7.

[10] G. Zhou, Z. Zeng, J. X. Huang, and T. He, Transfer Learning for Cross-Lingual Sentiment Classification with Weakly Shared Deep Neural Networks, Proc. 39th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. – SIGIR 16, pp. 245254, 2016.

[11] P. Vateekul and T. Koomsubha, "A study of sentiment analysis using deep learning techniques on Thai Twitter data," 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016, pp. 1-6, DOI: 10.1109/JCSSE.2016.7748849.

[12] R. Ghosh, K. Ravi, and V. Ravi, "A novel deep learning architecture for sentiment classification," 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), 2016, pp. 511-516, DOI: 10.1109/RAIT.2016.7507953.