



A Survey on Sentiment Analysis

Ms. Adagale Sushadevi Shamrao¹, Mr. Govind Hanumandas Rathi²,

¹Computer Engineering ,Trinity Academy of Engineering ,Pune India

²Computer Engineering ,Trinity Academy of Engineering ,Pune India

ABSTRACT

Sentiment Analysis, also known to everyone as Opinion Mining is a task of extracting and analyzing people's evaluations, opinions, sentiments, perceptions, attitudes, etc. having a wide range of entities like subjects, items, businesses, and services. The enormous growth in internet-based applications like websites, social-networks, and blog sites leading people to generate high amount of data in form of their opinion and reviews about the product, services and many things that they use in their day to day life. Sentiment analysis has placed in market as a powerful tool to analyze the public moods and views on which business get insight about their product. The sentiment analysis as a tool is used by businesses, governments, analysts and researchers to analyze the opinions of people. This paper gives a complete overview of sentiment analysis strategies, trends and challenges in form of a survey. The paper presents the application of sentiment analysis, approaches used for it and also discusses about challenges to be overcome to make clear future directions.

Keywords- *Sentiment Analysis*, NLP, AI,

1. INTRODUCTION

Sentiment Analysis is a project of natural Language Processing (NLP) that goals to extract sentiments and critiques from texts. Except, new sentiment analysis techniques begin to include the statistics from text and different modalities such as visual statistics. This research topic is conjoined below the sphere along emotion reputation. According to, affective computing and sentiment evaluation are the keys to the improvement of synthetic Intelligence (AI). Moreover, they've a first-rate capacity whilst applied to diverse domain names or structures. The challenge of sentiment evaluation can be taken into consideration as a textual content class trouble because the method consists of numerous operations that emerge as with classifying whether a given textual content expresses an effective or bad sentiment. However, sentiment evaluation can also appear an easy process, however in reality, it calls for deliberating many NLP subtasks like sarcasm and subjectivity detection. Moreover, the textual content is not usually organized as in the books or newspapers and might include many orthographic mistakes, idiomatic expressions, or abbreviations.in recent times, sentiment analysis has become well acknowledged, now not handiest among researchers, however also companies, governments, and groups. The growing use of the internet have made the web end up the established and the maximum crucial source of



data. Millions of human beings express their evaluations, and sentiments in forums, blogs, wikis, social networks, and other web assets. The ones opinions and sentiments are very applicable to our daily lives, and hence there may be a need to analyze this user-generated information so that it will robotically reveal the public opinion and help decision-making. As an example, Twitter posts have been used to expect election results. For this reason, the sector of sentiment evaluation gained more hobby within the closing one and a half of many years among research groups. Considering 2004, sentiment analysis has end up the fastest growing and the maximum active studies place, as there has been a massive boom in the range of papers specializing in sentiment analysis and opinion mining these days.

To the exceptional of our expertise, the present surveys often do now not include most of the people of sentiment analysis techniques and pay attention only on some supervised system getting to know and lexicon-based totally strategies. Although this paintings has additionally discussed these methods, it differs from the previous research with the aid of covering the most used strategies. further to that, different surveys investigate sentiment analysis from particular points of view which include demanding situations or consciousness on particular domains together with film reviews. This paper presents a more comprehensive study of sentiment analysis because it discusses this field from extraordinary points of view, as it consists of many research parts associated with sentiment analysis which includes challenges applications equipment, techniques, etc. this is very helpful for researchers and learners as they can find an widespread amount of information approximately this subject in one paper. Our paper differs from other surveys additionally with the aid of providing specified benefits and disadvantages of sentiment analysis techniques, which may assist researchers to pick the best technique to their issues. The giant contributions of this survey may be summarized as:

- A massive range of literatures has been studied to explain the sentiment analysis process in detail and pick out the tools to carry out this task.
- Categorizing the most used sentiment analysis techniques and summarizing them in quick info to have an overview of available techniques
- Comparisons of available processes to pick out the proper one for a given utility.
- Summarizing sentiment analysis applications and challenges to display the brand new trending researches.

2. TYPES OF SENTIMENT ANALYSIS

People have a huge variety of emotions – unhappy or glad, interested or bored to death, and fine or negative. Exceptional sentiment analysis models are available to seize this kind of feelings. To apprehend a way to observe sentiment evaluation in the context of your enterprise operation - you need to understand its different sorts.

2.1 Fine-Grained

This sentiment analysis model allows you derive polarity precision. You can behavior a sentiment analysis across the following polarity classes: very fine, superb, impartial, poor, or very bad. Nice-grained sentiment evaluation is beneficial for observe of opinions and rankings. You may get unique results in terms of the polarity



of the input. But, the technique to recognize this may be more labor and price-intensive as compared to other sorts.

For example- The rating system ranging from 1 to 5 where 1 is considered as negative and 5 is considered as positive.

2.2 Emotion Detection

Emotion detection facilitates you stumble on feelings. This could include anger, unhappiness, happiness, frustration, worry, fear, panic, etc. Emotion detection systems typically use lexicons – a collection of words that deliver sure emotions. Some superior classifiers additionally make use of sturdy gadget gaining knowledge of (ML) algorithms. It's encouraged to use ML over lexicons because human beings express feelings in a myriad of approaches. The benefit of the use of this is that an organization can also recognize why a customer feels a specific way. This is more algorithm-based totally and might be complex to understand in the beginning. The example can be taken as a review "The product is going to kill me" is an emotion of fear.

2.3 Aspect-Based

Its purpose is to perceive an opinion regarding a selected detail of the product. As an instance, the brightness of the flashlight within the telephone. The thing-primarily based analysis is typically utilized in product analytics to hold a watch on how the product is perceived and what are the sturdy and susceptible points from the purchaser's factor of view. The review of any product can be taken as an example where its different attributes are aspects of the product.

2.4 Intent Analysis

Its purpose is to determine what sort of aim is expressed inside the message. It's far typically utilized in customer support structures to streamline the workflow. This is a deeper knowledge of the goal of the customer.

3. APPLICATIONS OF SENTIMENT ANALYSIS

Sentiment analysis could be very useful in a wide range of software domain names starting from figuring out customer opinion to monitoring the intellectual health based on patient's social media posts. in addition to this, the emergence of recent technology which includes big information, Cloud Computing ,and Block chain has widened the location of packages supplying for sentiment evaluation unlimited opportunities to be carried out in almost each area. As an example, some of the common software domain names of sentiment analysis are defined within the following subsections.

3.1 Business Intelligence

Using sentiment analysis inside the area of business intelligence has many benefits, as an instance, groups can exploit the results of sentiment evaluation to make product enhancements, study the client's remarks, or undertake a brand new advertising strategy. Analyzing customers' perceptions of services or products is the most common software of sentiment analysis in the vicinity of commercial enterprise intelligence. However, these analyses aren't applicable most effective to product producers, however costumers can make use of them additionally to compare products and make a better choice. They analyzed the opinions using NRC emotion lexicon that classified customers' critiques into eight feelings (anger, fear, trust, anticipation, disappointment,



surprise, disgust, and joy) and sentiments (high quality and poor). Their results show that sentiment evaluation can assist to perceive the customers' behaviors and triumph over dangers to satisfy the clients' delight.

3.2 Healthcare System

The software of sentiment evaluation in the medical area has received so much interest these days. This utility allows healthcare actors to obtain information about the sicknesses, ad-verse drug reactions, epidemics, and patients' mood, and examine them to offer higher healthcare services. But, it is difficult to apply sentiment evaluation in the sort of domain because of a few faced problems like terminology as determined inside the work of et al. Clark et al. identified and analyzed tweets associated with the patient reveal in as an extra informative device for tracking public fitness. They amassed about 5.3 million breast cancer associated tweets the usage of Twitter's public streaming API for over a length of 12 months. After pre-processing, they analyzed tweets the usage of a logistic regression classifier and a convolutional neural community model to sift tweets relevant to breast cancer affected person stories. The authors founded that positive experiences were shared concerning patient remedy, elevating guide, and spreading focus. This paintings proves that social media can offer a wonderful outlet for sufferers to discuss their needs and issues. As a result, reading patient's generated data on social media using sentiment evaluation could be very beneficial to infer patient healthcare coverage and pick out his remedy needs.

3.3 Recommendation System

A recommender system is a set of rules that objectives to suggest relevant items (films, track, or product to shop for) to users. An efficient recommender system can generate a big quantity of profits for a few industries. For that reason, such systems can enjoy the software of sentiment analysis to make a higher recommendation. Within the paintings of Li et al., the authors proposed KBridge; a smart film advice device the use of sentiment analysis of microblogs. The machine identifies dialogue groups in microblogs which are correlated with a given topic and use a proposed novel sentiment-conscious association rule mining algorithm to investigate the correlation between groups. This research employs the feelings expressed in micro blogs to identify frequent program patterns and deduce the association rule of film/television software.

4. APPROACH TO SENTIMENT ANALYSIS

Sentiment analysis is a lively and flourishing studies field and can be applied in many domains. For this reason, researchers propose, evaluate, and compare exclusive tactics constantly. The aim is to increase the performance of sentiment analysis and to find answers to this subject challenges. Furthermore, applying sentiment analysis in new domain is a great incentive and makes this venture extra critical. But, choosing the appropriate approach for sentiment analysis could be very essential and essential. The purpose of this section is to provide an overview of the most used processes to perform sentiment analysis.

4.1 Pre-Processing

Pre-processing is a process used to extract actual required data from datasets. The pre-processing phase consist of data collection, data pre-processing.

4.1.1 Data Collection



Step one in sentiment evaluation is to have textual statistics, numerous sources and plenty of tools are available, a good way to obtain it. In widespread, textual content information can be created or collected as a part of research or through web scraping and crawling. Therefore, enhancing textual data with other types of facts (e.g., telecoms records, geospatial information, and video records) to perform sentiment analysis can cause thrilling results. There are multiple data sources like social networks, websites, blogs, reviews, forums, etc. The input data is generally in form of CSV files, JSON or XML files, sometimes it may be media files.

4.1.2 Data Pre-processing

The records received from extraordinary resources particularly social media are normally unstructured. The uncooked shape of those facts may additionally incorporate a number of noise and all types of spelling and grammatical mistakes. Consequently, it's far important to clean and preprocess text earlier than analysis. The aim of the preprocessing step is not handiest to get higher analysis however additionally to reduce the dimensionality of input data since many words are useless and should be removed, they do not have any impact on the text polarity (e.g., articles, prepositions, punctuation, unique characters). The process includes different tasks like Tokenization, stop-word removal, Lemmatization.

4.2 Machine Learning Approach

Machine studying is the most broadly-used method. It is based on system getting to know algorithms and linguistic features to carry out sentiment type. The lexicon-primarily based method makes use of sentiment lexicon which represents a list of phrases and terms which are usually used to explicit high-quality or negative sentiments. However, hybrid strategies combine device studying and lexicon-based totally procedures to enhance sentiment analysis performance.

4.2.1 Supervised Learning Approach

Supervised techniques require categorized schooling documents, wherein the labels are typically the training (e.g., nice, impartial, and bad). There are four styles of supervised category strategies which might be linear, probabilistic, rule-based, and decision tree. Inside the following subsections, a quick explanation and contrast of the most supervised class procedures typically used for sentiment evaluation.

4.2.2 Unsupervised Learning Approach

From time to time, its miles hard to accumulate and create classified datasets, specifically for textual records which is unstructured maximum of the time. That is because their generation requires people labeling data that is too labor-intensive and time-eating .on the other hand, its miles easier to accumulate unlabeled datasets and then, classify those using unsupervised studying tactics. Those techniques employ the files' statistical properties inclusive of word co-prevalence, NLP processes, and current lexicons with emotional (or) polarized words. However, in device mastering, unsupervised methods in the discipline of sentiment evaluation typically use clustering, that can classify facts into one-of-a-kind categories without specifying precisely which sentiment is represented by using each class. In other phrases, the clustering technique divides records into businesses (clusters), wherein the information of a cluster are very similar from a particular point of view than the records of various clusters.



4.2.3 Semi-Supervised Learning Approach

Semi-supervised studying(SSL) procedures also are used when there are difficulties in acquiring categorized statistics, however not like unsupervised techniques, this approach uses a small set of initial labeled training records to guide the feature studying system. Consequently, it suits in among supervised and unsupervised procedures. SSL approaches make complete use of amounts of low-price unlabeled data, save a lot of time and effort, and gain a classifier with strong generalization capacity in addition to greater categorized statistics

4.3 Lexicon-based Approach

Lexicon-based (also called understanding based totally) approach is one of the two main processes used for sentiment analysis and requires a lexical useful resource named opinion lexicon(a predefined list of words) which associates word to their semantic orientations terrible or positive words the use of ratings. A rating can be as an instance an easy polarity value such as +1, -1 or zero for positive, bad, or neutral phrases respectively, or a value reflecting the sentiment energy or intensity. The very last orientation of a document is obtained by calculating the semantic orientation values of the words that compose it. A document is tokenized into single phrases or micro terms, and then sentiment values from the lexicon are assigned to each element. To conclude the overall sentiment of a given report, components, or set of rules (e.g., sum and average) can be implemented. The lexicon based approach may be manual approach or dictionary based approach or corpus-based approach which uses statistics or semantics.

4.4 Hybrid Approach

The hybrid approach combines both lexicon and machine learning approaches. It combines the through put of lexical analysis with the ability of system learning approaches to cope with ambiguity and combine the context of sentiment words. The main reason behind the hybrid technique is to inherit high accuracy from system gaining knowledge of and balance from lexicon-primarily based method. The hybrid method combines strategies from the two previous techniques so as to triumph over their boundaries and take advantage of their advantages. For that, the lexicon approach scores are applied as enter functions to the sentiment classifier. For that reason, sentiment lexica play a critical position in the hybrid approach which is commonly known to reap a better overall performance.

5. CHALLENGES

The sentiment analysis has multiple challenges to tackle. The major challenges which come in sentiment analysis are sarcasm in the text, negation handling, detecting the spam, anaphora and reference resolution. The major challenge is low-resource availability of many languages. The code-mixed data is also a large problem in sentiment analysis. Many times people just give their opinion for spamming or sarcastically which impacts on evaluating the sentiments from those opinions. Handling of the negatives like neither, nor, not as important as single mistake can impact overall polarity of opinion. Sometimes many terms have same meaning which may impact on what the opinion is given. Overall these are impacting on performance of sentiment analysis and need to be tackled.



6. CONCLUSION

The paper presents an overview of sentiment analysis. The purpose of this paper was to discuss different approaches, challenges and applications of sentiment analysis. The paper categorizes different applications and approaches and how they are used in day to day life.

7. REFERENCES

- [1] B. Liu, Sentiment Analysis, 2015, pp. 1–367, <https://doi.org/10.1017/CBO9781139084789>.
- [2] I. Chaturvedi, E. Cambria, R.E. Welsch, F. Herrera, Distinguishing between facts and opinions for sentiment analysis: Survey and challenges, *Inf.Fusion*. 44 (2018) 65–77
- [3] E. Cambria, Affective computing and sentiment analysis, *IEEE Intell. Syst.*31 (2016) 102–107, <https://doi.org/10.1109/MIS.2016.31>.
- [4] Birjali, A. Beni-Hssane, M. Erritali, Analysing social media through big data using info sphere big insights and apache flume, in: *Procedia Comput.Sci*, 2017, pp. 1–6
- [5] F.J. Ramírez-Tinoco, G. Alor-Hernández, J.L. Sánchez-Cervantes, M. delP. Salas-Zárate, R. Valencia-García, Use of Sentiment Analysis Techniques in Healthcare Domain, 2019, pp. 189–212
- [6] S.M. Jiménez-Zafra, M.T. Martín-Valdivia, M.D. Molina-González, L.A.Ureña López, How do we talk about doctors and drugs Sentiment analysis in forums expressing opinions for medical domain, *Artif. Intell.Med.* 93 (2019) 50–57
- [7] Dzisevič, D. Šešok, Text classification using different feature extraction approaches, in: 2019 Open Conf. Electra. *Electron. Inf. Sci*, 2019, pp. 1–4
- [8] Yadav, D.K. Vishwakarma, Sentiment analysis using deep learning architectures: a review, *Artif. Intell. Rev.* (2019) 1–51
- [9] M.V. Mäntylä, D. Graziotin, M. Kuutila, The evolution of sentiment analysis—A review of research topics, venues, and top cited papers, *Comput. Sci. Rev.* 27 (2018) 16–32, <https://doi.org/10.1016/j.cosrev.2017.10.002>.
- [10] B. Liu, L. Zhang, A survey of opinion mining and sentiment analysis, in: *Min. Text Data*, Springer, US, Boston, MA, 2012, pp. 415–463, https://doi.org/10.1007/978-1-4614-3223-4_13.
- [11] R. Piryani, D. Madhavi, V.K. Singh, Analytical mapping of opinion mining and sentiment analysis research during 2000–2015, *Inf. Process. Manag.* 53 (2017) 122–150, <https://doi.org/10.1016/j.ipm.2016.07.001>.
- [12] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, *Ain Shams Eng. J.* 5 (2014) 1093–1113, <https://doi.org/10.1016/j.asej.2014.04.011>.
- [13] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, *Knowledge-Based Syst.* 89 (2015) 14–46, <https://doi.org/10.1016/j.knsys.2015.06.015>.
- [14] L. Zhang, S. Wang, B. Liu, Deep learning for sentiment analysis: A survey, *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 8 (2018) 1–25, <https://doi.org/10.1002/widm.1253>.
- [15] F. Hemmatian, M.K. Sohrabi, A survey on classification techniques for opinion mining and sentiment analysis, *Artif. Intell. Rev.* 52 (2019) 1495–1545, <https://doi.org/10.1007/s10462-017-9599-6>.



- [16] S. Rajalakshmi, S. Asha, N. Pazhaniraja, A comprehensive survey on sentiment analysis, in: 2017 Fourth Int. Conf. Signal Process. Commun. Netw, IEEE, 2017, pp. 1–5, <https://doi.org/10.1109/ICSCN.2017.8085673>.
- [17] B. Batrinca, P.C. Treleaven, Social media analytics: a survey of techniques, tools and platforms, *AI Soc.* 30 (2015) 89–116, <https://doi.org/10.1007/s00146-014-0549-4>.
- [18] V.A. Kharde, P.S. Sonawane, Sentiment Analysis of Twitter Data: A Survey of Techniques, 2016, pp. 5–15, <https://doi.org/10.5120/ijca2016908625>. [28] P. Yang, Y. Chen, A survey on sentiment analysis by using machine learning methods, in: 2017 IEEE 2nd Inf. Technol. Networking, Electron. Autom. Control Conf, IEEE, 2017, pp. 117–121, <https://doi.org/10.1109/ITNEC.2017.8284920>.
- [19] H. Thakkar, D. Patel, Approaches for Sentiment Analysis on Twitter: A State-of-Art Study, 2015, pp. 1–8, <http://arxiv.org/abs/1512.01043>. [30] R. Liu, Y. Shi, C. Ji, M. Jia, A survey of sentiment analysis based on transfer learning, *IEEE Access.* 7 (2019) 85401–85412, <https://doi.org/10.1109/ACCESS.2019.2925059>.
- [20] N.F.F. Da Silva, L.F.S. Coletta, E.R. Hruschka, A survey and comparative study of tweet sentiment analysis via semi-supervised learning, *ACM Comput. Surv.* 49 (2016) 1–26, <https://doi.org/10.1145/2932708>.
- [21] H.H. Do, P. Prasad, A. Maag, A. Alsadoon, Deep learning for aspect-based sentiment analysis: A comparative review, *Expert Syst. Appl.* 118 (2019) 272–299, <https://doi.org/10.1016/j.eswa.2018.10.003>.
- [22] C.C. Aggarwal, Machine learning for text, *Mach. Learn. Text.* (2018) 1–493, <https://doi.org/10.1007/978-3-319-73531-3>.
- [23] S. Behdenna, F. Barigou, G. Belalem, Sentiment analysis at document level, in: SmartCom 2016, 2016, pp. 159–168, https://doi.org/10.1007/978-981-10-3433-6_20.
- [24] N. Indurkha, F.J. Damerau, Handbook of Natural Language Processing, 2010, pp. 1–704.
- [25] M. Tubishat, N. Idris, M.A.M. Abushariah, Implicit aspect extraction in sentiment analysis: Review, taxonomy, opportunities, and open challenges, *Inf. Process. Manag.* 54 (2018) 545–563, <https://doi.org/10.1016/j.ipm.2018.03.008>.
- [26] M.E. Mowlaei, M. Saniee Abadeh, H. Keshavarz, Aspect-based sentiment analysis using adaptive aspect-based lexicons, *Expert Syst. Appl.* 148 (2020) 1–13, <https://doi.org/10.1016/j.eswa.2020.113234>.
- [27] L. Mai, B. Le, Joint sentence and aspect-level sentiment analysis of product comments, *Ann. Oper. Res.* (2020) <https://doi.org/10.1007/s10479-020-03534-7>.
- [28] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding, 2018, pp. 1–16, <http://arxiv.org/abs/1810.04805>.
- [29] B. Liu, Sentiment analysis and opinion mining, *Synth. Lect. Hum. Lang. Technol.* 5 (2012) 1–167, <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>.
- [30] T. Chen, R. Xu, Y. He, X. Wang, Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN, *Expert Syst. Appl.* 72 (2017) 221–230, <https://doi.org/10.1016/j.eswa.2016.10.065>.