

A Review on Sentimental Analysis Using Machine Learning Techniques

Mr.Nitin Gupta¹, Prof. Vijay Prakash Sharma²

^{1,2}Dept of Computer Science & Engineering

^{1,2}Shri Ram college of Engineering and Management,Banmore

Abstract- For businesses and manufacturers, opinion information is extremely essential. They frequently want to know intimately what their goods and services think of customers and the public. It is nonetheless unrealistic to read every article manually on the site and extract valuable views from it. If you do it manually, too much info is available. Sentiment analysis provides for efficient and cost-effective large-scale processing of data. To learn more about sentiment analysis, this author examines how sentiment analysis is used by businesses to identify their strengths & limitations. This paper summarizes the sentiment analysis on amazon reviews with its application and classification levels. This work reviews sentiment classification techniques and also studies feature selection in sentiment analysis. In addition, it provides the introduction of natural language processing and its techniques, and limitations. At last, we described text mining in this paper.

Keywords- Data Mining, Opinion Mining, Sentiment Analysis, Natural Language Processing, Classification Techniques, Machine Learning, Text Mining.

I. INTRODUCTION

Data mining is the research phase of "document data discovery." This is a technique for making decisions and extracting information from huge amounts of data. It is a technique for removing knowledge from data. Sentiment analysis is a term that relates to the processing of natural language [1]. Sentiment Analysis (SA), sometimes referred to as Opinion Mining (OM), is a study of people's sentiments, views, & emotions about various entities. Opinions are important to virtually all human endeavors because they have a significant effect on how people behave. When a decision must be made, people seek out the views of others. In the real world, businesses and organizations are interested in learning what the majority think about their goods & services. On the other side, customers need to identify what others think of a product before making a purchase. Individuals have historically sought counsel from friends and family, while corporations have conducted surveys or organized focus groups. Nonetheless, as social networks like Facebook & Twitter have grown in popularity, people and organizations

have begun to rely on the data given by these platforms to aid in their decision-making processes. SA, often called opinion mining, developed out of this context [2].

Sentiments convey the customer's perspective, which may be good or negative, or it may be neutral [3]. Sentiment classification also referred to as OM, is a method of determining the sentiment category to which textual content belongs automatically. We may broadly classify their views, comments, and documents into two categories. There are two types of sentiment: numerical and categorical. A common instance of numerical sentiment is the rating system seen on e-commerce websites. The business evaluates the public's response by using this grading system. Categorical SA is a method utilized to categorize comments & evaluations. Binary (positive & negative), ternary (positive, negative, & neutral), as well as numerous categories (happy, sad, angry, etc.), are available [4]. Opinions have evolved into critical components of all human endeavours.

One of the main methods to promote SA is NLP (Natural Language Processing). It is an artificial intelligence sub-domain that enablesthe computer-aided system to comprehend many languages as spoken or spoken by human beings. The main aim of the NLP is to comprehend the unstructured materials in social media and to organize the same for SA. NLP emphasizes primarily reading & interpreting the text of the free form in an analysis format. Opinion mining is therefore the main application field in which NLP is a major component. Another frequent application where NLP plays a key role also helps search engines like Google to adjust their search algorithms to comprehend and interpret various situations and to generate appropriate search results.[5].

Information Retrieval (IR) plays an essential function in online search engines in accessing the most relevant user input query information. It is a mainstream of online search engines and their fundamentals. Recovery of information is how information is collected and presented from the biggest information collection according to the user's input query. Whenever a user wants to access the data, a formal statement must be entered in a search engine. This formal statement is

also known as the input query of a search engine. A query does not receive and display a single information resource in the biggest information resource collection. Instead, numerous information sites matched by the input query are displayed. The user will be presented with the most related information resources.

The remaining paper is systematized as follows: section II gives a small introduction to opinion mining. Section III elaborates a sentiment analysis with its applications and classifications levels. It talks about various sentiment classification techniques in Section IV. Section V presents a small part of feature selection in sentiment analysis. The overview of NLP (Natural language processing) has given with these applications and limitations in Section VI. Section VII provides a detailed overview of text mining also describes its techniques, applications, and limitations. Section VIII presents the literature review. At last, Section IX concludes this work.

II. OPINION MINING

Opinion mining (OM) is a technique for identifying and extracting subjective data from text documents. Sentimental analysis is another term for opinion mining. The sentiment may be based on his or her assessment of goods customer reviews.

Two techniques exist for predicting sentiment:

- 1) **Direct opinions:** Text documents that provide an explicit favourable or negative opinion on a product. For instance, "This mobile's battery backup is inadequate."
- 2) **Comparison Opinions:** Opinions expressed in a text document to compare one object to another object.

A. Opinion Mining Techniques

Techniques are classified into three broad categories (fig. 2):

- **Supervised Learning Techniques:** Support Vector Machines (SVM), Neural networks (NNs), Multi-Layer Perceptron (MLP), Naive Bayes (NB) Classification, Decision trees (DTs), and Maximum Entropy (MaxEnt) are the most frequently utilized supervised learning methods.
- **Unsupervised Learning Techniques:** The most often utilized techniques are clustering, expectation-maximization, matrix factorization, and principal component analysis (PCA).
- **Case-Based Reasoning:** It is a newly discovered artificial method. CBR is a sophisticated technique

for computer reasoning that solves problems in real-time. The solution is maintained in a CBR repository, often referred to as a case base.

B. Applications

The following are the most common applications of SA & OM.

- **Opinion spam detection:** People may write product reviews with negative intent. Opinion mining and sentiment analysis can distinguish between spam and non-spam content in these evaluations.
- **Purchasing Product or Service:** People may readily assess other people's opinions and experiences regarding any product or service using this approach, as well as compare rival brands.
- **Quality Improvement in Product or service:** This allows the manufacturer to collect both critical and positive feedback on their product or service, allowing them to enhance the quality of their product or service.
- **Marketing research:** It is possible to evaluate a product or service in light of a new government policy. All of these findings may be credited to collective intelligence research.
- **Policy-Making:** Policymakers may use sentiment analysis to ascertain citizens' perceptions of a certain policy and use this information to develop new, more citizen-friendly policies.
- **Decision Making:** Individuals' perspectives and experiences are very valuable in the decision-making process. It provides analyzed public opinion that may be utilized to make sound decisions.

III. SENTIMENT ANALYSIS

Sentiment Analysis (SA) has become numerous areas of computational study in NLP [6], [7]. Generally, SA is concerned with the mining of information of sentiments or views of a group on a certain subject. Furthermore, at the document level sentiments are obtained from certain applications. For additional research investigations including sentiment or opinion-based studies, opinion-based summarization, emotion or mood extraction, and genre differences are taken into consideration.



Fig. 1: Sentiment Analysis

Generally, SA entails series of intricate processes. Some tasks surround the study, including subjective analysis, sentiment classification, object or aspect-based extraction, & opinion holder extraction. Subjective analysis entails analyzing text document or phrase to determine if it is subjective or objective. Those documents and phrases labeled as aims are immediately deleted after this stage since they are not of much value for the SA process. Sentiment classification entails determining the polarity of filtered sentences' sentiments. Based on the circumstances, these phrases are classified as positive, negative, or neutral sentiments. A critical task is an aspect or object-based extraction, which is the primary focus element in SA. In certain instances, the task of identifying the opinion holder becomes critical since it is necessary to know the author of view [8].

A. Application of Sentiment Analysis

SA has a broad variety of applications. SA may be employed in a variety of circumstances, including those involving politics, product evaluations, movie reviews, and posts on social media. Figure 3 illustrates how SA is used in the context of movie reviews. From a wider & more global perspective, SA has been pushed into the spotlight due to the possibility of gathering more accurate and comprehensive data from social media sites such as Facebook, LinkedIn, & Twitter [9].

Different application	Different rating
Movie review	★★★★★
Product review	☑️☑️☑️☑️
Politics	👍🗨️
Public sentiment	😊😞
Social sites	🌐🐦📘📷📺

Fig. 2: Application of sentiment analysis

B. Classification Levels

SA is seen as a three-layered strategy. 1st layer is a document-based layer. 2nd is a sentence-based approach. 3rd level is the aspect level, which is sometimes referred to as the word or phrase level.

- 1) **Document Level:** 1st level is referred to as document level in SA. At this level, the whole of a document is evaluated for SA purposes. In this approach of performing SA, the source of the opinion is assumed to be a single source or a single entity [10]. At this level, sentiment regression has been found to constitute a distinct feature of SA [11], [12]. Numerous studies have resorted to supervised learning to estimate document evaluations to ascertain the degree of a positive or negative perspective [13].
- 2) **Sentence Level:** Numerous research papers exist that categorize and analyze each phrase in a document or part of the text as subjective or objective. In one instance, the authors [11] suggest doing SA only on subjective phrases after classification. Machine learning has been a boon to academics and scholars in this area as a technique for identifying subjective sentences. In one research [12], researchers presented a model depending upon logarithmic probability rates & no. of the root, words to serve as a foundation for a scorecard used to categorize every subjective phrase categorized. Another article [13] proposed a model that incorporates sentiments of all words in a phrase to provide complete sentiment for the sentence under examination.
- 3) **Aspect Level:** Aspect Level SA is intended to solve limitations of SA at document & sentence levels. With the assistance of SA, fine-grained control may

be exercised. The objective of aspect level SA is to critically & analyze opinion. SA at the aspect level implies that an opinion may only be represented in one of four ways: positively, neutrally, negatively, or as an explicit objective sentiment [14][15].

IV. SENTIMENT CLASSIFICATION TECHNIQUES

Sentiment Classification techniques are classified as lexicon-based, machine learning, & hybrid approach.

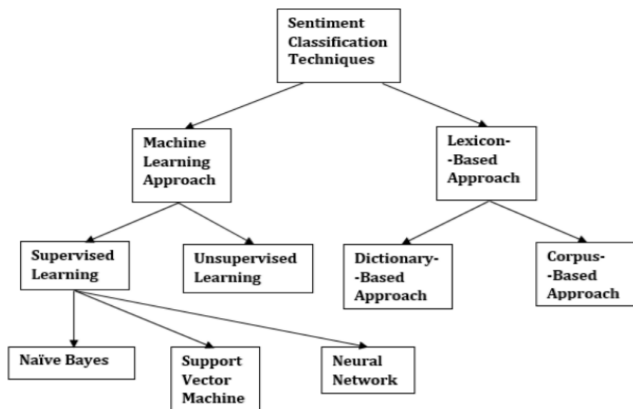


Fig. 3: Tree diagram of sentiment classification techniques

A. Machine Learning Approach

SA is a challenge of universal text classification that requires the incorporation of syntactic & linguistic characteristics. The text classification problem is defined as follows: We have a collection of training documents $Doc = \{Y_1, Y_2, \dots, Y_n\}$, each of which is labeled with class. The classification model is suitable for underlying record's characteristics that need to be classified according to the class label. Model is then utilized to predict the class label for a given instance of an undiscovered class. The difficult categorization issue occurs when an instance is assigned just one label. When the probabilistic value of labels is assigned to an instance, this is referred to as the soft classification issue.

1) Supervised Learning

The availability of labeled training data is required for the supervised methods. There are many types of supervised classifiers; however, NB(Naive Bayes),SVM (Support Vector Machine), and RF(Random Forest) are most commonly employed in SA.

- **Naïve Bayes (NB):** NB is a simple and easy but powerful category set of rules. It is primarily used for file-stage classification. Naive Bayes classifiers are computationally rapid whilst taking selections. Naïve

Bayes Classifier is well used in real problems together with electronic mail Spam detection, Sentiment analysis, and Sexual content detection.

- **Support Vector Machine (SVM):**SVM is a discriminative classifier i.e. popular due to its successful text categorization method. It is a statistical classification technique SVM maps input (actual-valued) feature vectors right into a higher 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 on the minimization of the empirical chance, that's a try and optimize the performance of the gaining knowledge of set.
- **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 with 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.

2) Unsupervised Learning

This is no longer a category, but they do not offer its proper goals in any way, necessitating the use of clustering.

Table 1: Comparing Techniques of Sentiment Classification

Sentiment classification technique	Effectiveness	Accuracy	Advantages	Limitations
Supervised learning approach	Slow	Very High	-Allow analysis of numerous categories -Efficient discovery of subjectivity issue -Less affection by noisy data	-Dependence issue with labelled training documents -Requires the presence of human effort and linguistic knowledge -Higher cost, Require more to train specially for high dimensional data -Rely on human participation
Semi-supervised learning approach	Medium	High	-Good performance for ambiguous data reviews -Higher accuracy can be achieved using less human annotation effort possible in moderate amount of time	-Noise in the unlabelled samples makes classification troubles
Unsupervised learning approach	Fast	Medium	-It does not require much human participation -It needs to be efficient and widely applicable	-Its ability to analyse multiple categories has not still been proven -Get easily affected by noisy terms -The number of clusters in most cases is unknown -The accuracy can sometimes be relatively low with unstable results

B. *Lexicon-based Approach*

Numerous emotion class assignments include the usage of opinion expressions. Positive opinion phrases are used to describe certain preferred activities, whereas negative opinion phrases are used to describe a few undesirable occurrences. Additionally, there have been opinion expressions as well as idioms that contribute to this opinion lexicon's composition. There are three critical ways of delivering or compiling an opinion phrase list. This manual technique is time-consuming but is not utilized alone. It has been used in combination with those other automated processes as a last check to ensure that automated techniques do not make errors.

1) **Dictionary-based Approach**

This little collection of opinions related to well-known guidelines has been gathered manually. This collection is then supplemented by the search for synonyms and antonyms in famous companies such as WordNet or the Merriam-Webster dictionary. These additional words were added to the basic list, as well as the following iteration begins evolving. When no new word is identified, this iterative mechanism comes to a halt. Once the procedure is complete, this manual inspection may be used to identify and fix problems. This dictionary-based method's main shortcoming has been its inability to detect opinion words that discipline-but also context-specific orientations.

2) **Corpus-based approach**

This corpus-based method alleviates difficulties associated with learning opinion words when they are used in context-specific contexts. Its methods are based on syntactic patterns and styles that often coexist with a single root list of opinion words to discover more opinion words in a big corpus.

V. FEATURE SELECTION IN SENTIMENT ANALYSIS

This job of SA is classified as a sentiment classification issue. This SC issue begins with extraction as well as a selection of text characteristics. Several current features include the following: These terms presence, as well as frequency, are used interchangeably. Individual words or word n-grams, and their frequency counts, are included in features. It either weighs binary words (null if they exist, one if it does not) or it utilizes term frequency weights to show the relative significance of characteristics. Parts of speech (POS): Identifying adjectives that are significant indications of the point of view of a person. Opinion words or phrases: they are often used words or phrases that express opinions, whether

good or negative, liked or disliked. On the other hand, some sentences convey views without the use of opinion words. As an illustration, it cost me both an arm and a leg. Negative words: the existence of negative words have the potential to change an individual's viewpoint, since not good equals bad [16].

VI. NATURAL LANGUAGE PROCESSING

NLP is a subfield of AI (Artificial Intelligence) as well as Linguistics concerned with programming computers to comprehend statements or phrases written in human languages. NLP was created to facilitate user interaction as well as to fulfill that user's desire to speak with the computer in natural language. Given that not all users are fluent in machine-specific languages, NLP assists users who lack the time or inclination to acquire new languages to improve existing ones. A language may be thought of as a collection of rules or as a collection of symbols. Combinations of symbols have been used to communicate or broadcast information. These Rules tyrannise symbols. NLP is fundamentally divided into two components: Natural Language Understanding but instead Natural Language Generation, which also develops with job of text comprehension as well as generation [17].

A. *Applications of NLP*

Text-based applications These applications include searching over a certain topic or phrase in such a database, extracting data from a long text, translating one language to another, as well as summarizing information for a variety of purposes.

Exchange-based applications Typical instances include frameworks for question-answering, administrations that can be given so over the phone without assistance of an operator, instructional systems, voice-controlled machines (that accept instructions through speech), as well as frameworks for general problem-solving [18].

B. *Limitations of NLP*

In principle, natural language processing is a very appealing way of human-computer interaction. Natural language recognition requires a thorough understanding of the external environment. Earlier systems, such as SHRDLU, used limited "block words" but also vocabularies and functioned very well. However, it quickly failed when the systems were expanded to more realistic scenarios including ambiguity as well as complexity seen in the real world. Thus, creating an NLP application is difficult since computers need people to "talk" to them in a precise, clear, and highly organized

programming language. Natural language processing is plagued by ambiguity, vagueness, as well as uncertainty. These issues must be addressed regardless of whether one is working with a single phrase or a speech. That key issue in NLP is ambiguity. Ambiguity is a term that refers to an expression (word/phrase/sentence) that has many meanings. Ambiguity may arise at many levels of analysis, including lexical, syntactic, semantic, discourse, as well as pragmatic [19].

VII. TEXT MINING

This process of extracting important information, knowledge, as well as patterns through unstructured text documents is known as text mining. Text classification, text clustering, document summarizing, as well as sentiment analysis, are examples of text mining tasks. There are a variety of methods as well as tools for mining text for useful information for future prediction and decision-making. Text mining has become an important study area. By extracting data from different assembled assets, Text Mining reveals fresh as well as previously hidden facts. A survey of Text Mining methods & applications is presented in this article. Unstructured text documents of different sources include a large quantity of data that should not be processed in any way to extract valuable information.

Text mining is a technique of extracting information from huge volumes of unstructured text to find new, previously undiscovered information and knowledge. In general, humans find unstructured language to be simple, while computer programmes find it to be very difficult [20].

A. Text Mining Techniques

- 1) **Information Extraction:** The first step for computers is to detect amorphous typescript by identifying keywords as transactions within the text. Its goal is to extract useful information from large amounts of text.
- 2) **Information Retrieval:** For a longer time, database systems have been under development in the area of information retrieval. The goal of information retrieval is to get a document that contains the exact information that the user has obtained.
- 3) **Summarization:** A content summary is a process of gathering as well as creating succinct 3 representations of documents containing original text. To summarize, that raw text is taken initially, followed by preparation and processing procedures. Three techniques are used in pre-processing, including tokenization, stemming, as well as word elimination.

- 4) **Clustering:** It is the division of a set of items or data into a collection of relevant but comprehensible subclasses. Clustering is mostly used to create a collection of documents and files that are similar. This clustering method divides records in a dataset into groups so that themes within each cluster were similar but topics across clusters are distinct.
- 5) **Categorization:** These main topics of a text were identified during classification. It is accomplished by categorizing these papers into a set of preset categories. This is possible to classify this text as a "bag of words." Because information extraction attempts to process that real information, while classification does not [21].

B. Text Mining Applications

- Extraction Transformation Loading
- Competitive Intelligence
- Customer Relationship Management (CRM)
- Human resource management
- Questioning in Natural Language
- Market Analysis (MA).

C. Limitations in Text Mining

The complication may be discovered early on in the in-text mining process. There is preprocessing step, many rules and standards were carefully established to govern its content, making text mining more efficient. Unstructured data must first be transformed into an intermediate format before pattern analysis can be performed, but this step with the in-text mining process will have its own set of disadvantages. That actual topic of data may be lost several times owing to a change within the text's sequence [22].

VIII. LITERATURE REVIEW

Some of the research work which guides me in the way of completing my paper is discussed below:

(Y. A. Solangi et al., 2018) Opinion mining has lately been a lively study area due to research difficulties. Natural Language Processing (NLP) methods for OM but also SA were discussed in this work. Initially, NLP is reviewed, followed by a discussion of the most frequent as well as effective preprocessing procedures. This work analyses as well as reviews opinion mining at different levels. Finally, problems were discovered as well as recommendations for opinion mining but also sentiment analysis were made [23].

(S. Naz et al., 2018) This suggested technique tackles the categorization of Twitter sentiments by utilizing an ML classification model that uses many textual features, notably Twitter data N-grams. Additionally, they examined the effect of weighting on classifier accuracy using three distinct weighting methods. Additionally, a sentiment score vector derived from tweets is utilized to augment this SVM classifier's performance [24].

(M. Wongkar et al., 2019) The study was aimed at creating a sentiment analysis app of presidential candidates from the Republic of Indonesia for 2019 utilizing the Python programming language. According to with findings of this research, the positive sentiment polarity of Jokowi-Ma'ruf Amin combination was 45.45%, while its Prabowo-Sandiaga duo got a positive sentiment score of 44.32 percent as well as a negative sentiment score of 55.68 percent. This combined data then was evaluated against training data in each presidential contender, yielding an accuracy of 80.90% 80.10%. Its naive Bayes, SVM, but also K-NN(K-Nearest Neighbor) techniques were compared in this research using RapidMiner. That naive Bayes precision value had been 75.58 percent, the SVM precision value has been 63.99 percent, as well as K-NN precision value has been 73.34 percent[25].

(Y. Jing et.al.,2019) Present work suggests an online review sentiment categorization technique dependent on text document subject analysis as well as semantic analysis of sentiment terms. To begin, the subject of text documents as well as a topic of words were determined using the LDA technique. Then, Word2Vec is used to train these word vectors upon those words in text documents as well as words within the sentiment dictionary. Semantic similarity is computed between words with documents, as well as among words in a text document & words in a sentiment dictionary. Lastly, utilizing these two types of semantic similarity, these words with the greatest similarity were chosen as the text document's main categorization characteristics. Experiments show that the system performs well but has the potential to enhance the impact of internet reviews on sentiment categorization [26].

(J. Rama krishnan et al.,2020) the objective of this work is to offer an overview of techniques as well as uses of opinion mining. This work provides an overview of OM as well as related techniques but also examines uses as well as challenges associated with OM based on prior research. With the development of internet technology, a vast amount of information is available just on the web for internet users. Those users not only browse to available resources upon that web but also provide comments, resulting in the creation of more valuable data. OM is considered a branch of NLP, information retrieval, and data mining. It is a method for

separating people's ideas but also observations from unformed material, that has become a more effective, appealing, as well as complicated problem like a result of the growth of online social media as well as a massive quantity of user input. Its basic concept of OM is categorizing text as positive (good), neutral (unexpected), or negative (bad) [27].

(G. Li, Q.et al.,2020) To allow a model to explicitly acquire sentiment knowledge included in Chinese text, this article presents a network model based upon sentiment information (SINM). For model components, they can use a Transformer encoder as well as LSTM. They can automatically detect sentiment knowledge in Chinese text using a Chinese emotional lexicon. SINM developed a hybrid task learning technique for purpose of learning useful emotional expressions as well as predicting sentiment trends. So, begin, SINM must acquire sentiment information from text. SINM will prioritize sentiment information above irrelevant information when supplementary emotional information is present. SINM outperforms as well as generalizes most current techniques in experiments conducted just on ChnSentiCorp and ChnFoodReviews datasets[28].

(K. Jia et.al.,2020) This work improves its technique for calculating the strength of emotions associated with various emotion categories within the emotion dictionary. To start, they utilize word2vec to train emotional word vectors from a broad microblog text, choose benchmark words for each emotional category and estimate their similarities among emotional words vector and benchmarking words vector, to gain the emotional intensity of distinct emotional classes. Then, they examine the effect of negative words as well as degree adverbs upon that emotional propensity of various categories of emotion words, as well as establish criteria for emotion semantic calculation. Finally, they integrate this emotion dictionary as well as semantic criteria to determine the emotional intensity of various emotional categories in microblogs, thus completing the emotion categorization. Our approach is successful, as shown by the experimental findings[29].

(A. Guezzaz et.al., 2021) The primary aim of this research is to construct and verify a heterogeneous categorization of traffic that can classify gathered events inside the networks. During the analysis and categorization of data activities, intrusion sensor systems suffer from various vulnerabilities. New techniques of analysis should be developed to build a suitable system to monitor circulating traffic to resolve this problem. The latest prototype is dependent on a suggested ML method comprising input layer, hidden layer, and output layer. Improvement of weights and the validation method for the prototype are suggested by a trustworthy training algorithm.

Before the analytical phase, preprocessing is performed to traffic gathered. This article discusses the validation mathematically of a novel ML categorization for heterogeneous traffic and detection of anomalies[30].

IX. CONCLUSION

A classification of sentiments is a difficult field with a great deal of fence since it includes the understanding of natural language & hidden emotions. It enables a wide variety of applications to profit from its results, including film reviews, product evaluations, news analytics and marketing, question answers, and knowledge bases. There are many fields in the classification of sentiments where much improvement using existing techniques is required. This study offers a short overview of the categorization of feelings, the kind of classification of sentiments, and the comparison of classification methods. People's interest in non-English languages in this classification of sentiments is increasing day after day as there remains a shortage of resources and studies on them. For many natural languages, building resources utilized in sentiment classification tasks are still required. The survey also reveals several significant difficulties in assessing sentiments.

REFERENCES

- [1] N. Supriya, V. Kallimani, S. Prakash, C. B. Akki, Twitter sentiment analysis using binary classification technique, Int. Conference on Nature of Computation and Communication, 2016, pp. 391–396.
- [2] Keith Norambuena, Brian & Lettura, Exequiel & Villegas, Claudio. (2019). Sentiment analysis and opinion mining applied to scientific paper reviews. *Intelligent Data Analysis*. 23. 191-214. 10.3233/IDA-173807.
- [3] K. T. Devendra, S. K. Yadav, Fast retrieval approach of sentimental analysis with implementation of bloom filter on Hadoop, International Conference on Computational Techniques in Information and Communication Technologies, 2016, pp. 529–551.
- [4] J. Steinberger, T. Brychcin, M. Konkol, Sentiment and social media analysis, Proc. 5th Workshop on Computational Approaches to Subjectivity, 2014.
- [5] Yi, S., Liu, X. Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. *Complex Intell. Syst.* 6, 621–634 (2020). <https://doi.org/10.1007/s40747-020-00155-2>
- [6] M. Bautin, L. Vijayarenu and S. Skiena "International Sentiment Analysis for News and Blogs". In ICWSM.
- [7] R. Prabowo and M. Thelwall." Sentiment analysis: A combined approach". *Journal of Informetrics* , 3(2): 143-157, b2009.
- [8] J. Brooke, M. Tofiloski and M. Taboada "Cross-Linguistic Sentiment Analysis: From English to Spanish". In RANLP, pp. 50-54.
- [9] Raktim Kumar Dey, Debabrata Sarddar, Indranil Sarkar, Rajesh Bose, Sandip Roy, "A Literature Survey On Sentiment Analysis Techniques Involving Social Media And Online Platforms", *International Journal Of Scientific & Technology Research* Volume 9, Issue 05, May 2020.
- [10] J. Kamps, M. Marx, R.J. Mokken, and M. De Rijke, —Using WordNet to measure semantic orientations of adjectives, In LREC , Lisbon, Portugal , Vol. 4, pp. 1115-1118, May 2004.
- [11] B. Liu, M. Hu, and J. Cheng, —Opinion observer: analyzing and comparing opinions on the web, In Proceedings of the 14th international conference on World Wide Web, pp. 342-351, May 2005, doi:10.1145/1060745.1060797.
- [12] L. Qu, G. Ifrim, and G. Weikum, , — The bag-of-opinions method for review rating prediction from sparse text patterns, In Proceedings of the 23rd international conference on computational linguistics, Coling 2010, Beijing , Association for Computational Linguistics, pp. 913-921, August 2010.
- [13] Y. Chen, and J. Xie, —Online consumer review: Word-of-mouth as a new element of marketing communication mix, *Management science*, Vol. 54, No. 3, pp. 477-491, March 2008, doi: 10.1287/mnsc.1070.0810.
- [14] Fang, and J. Zhan, —Sentiment analysis using product review data, *Journal of Big Data*, Vol. 2, No. 1, p.5, June 2015, doi: 10.1186/s40537-015-0015-2.
- [15] V. A. Kharde, and S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", *International Journal of Computer Applications*, Vol. 139, No. 11, pp. 5-15, April 2016, doi:10.5120/ijca2016908625.
- [16] Walaa Medhat, Ahmed Hassan, Hoda Korashy, Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal*, Volume 5, Issue 4, 2014, Pages 1093-1113, ISSN 2090-4479, <https://doi.org/10.1016/j.asej.2014.04.011>.
- [17] Khurana, Diksha & Koli, Aditya & Khatter, Kiran & Singh, Sukhdev. (2017). *Natural Language Processing: State of The Art, Current Trends and Challenges*.
- [18] Ponnusamy, Ramalingam. (2018). *A Systematic Survey of Natural Language Processing (NLP) Approaches in Different Systems*. 4.
- [19] Aiyasha Sadiya, Archana R Hegde, Anusha V Illur, Ashwini R, *International Journal of Engineering Research & Technology (IJERT)*, Vol. 7 Issue 01, January-2018.
- [20] Maheswari, M. (2017). *Text Mining: Survey on Techniques and Applications*. *International Journal of Science and Research (IJSR)*.

- [21] Tandel, Sayali & Jamadar, Abhishek & Dudugu, Siddharth. (2019). A Survey on Text Mining Techniques. 10.1109/ICACCS.2019 Vishal Gupta, Gurpreet S. Lehal, A Survey of Text Mining Techniques and Applications, Journal of Emerging Technologies In Web Intelligence, Vol. 1, No. 1, August 2009.
- [22] Tandel, Sayali & Jamadar, Abhishek & Dudugu, Siddharth. (2019). A Survey on Text Mining Techniques. 10.1109/ICACCS.2019.8728547.
- [23] Y. A. Solangi, Z. A. Solangi, S. Aarain, A. Abro, G. A. Mallah and A. Shah, "Review on Natural Language Processing (NLP) and Its Toolkits for Opinion Mining and Sentiment Analysis," 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2018, pp. 1-4, doi: 10.1109/ICETAS.2018.8629198.
- [24] S. Naz, A. Sharan and N. Malik, "Sentiment Classification on Twitter Data Using Support Vector Machine," 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2018, pp. 676-679, doi: 10.1109/WI.2018.00-13.
- [25] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985884.
- [26] Y. Jing, H. Gou, C. Fu and W. Sun, "Sentiment Classification of Online Reviews Based on LDA and Semantic Analysis of Sentimental Words," 2019 12th International Symposium on Computational Intelligence and Design (ISCID), 2019, pp. 249-252, doi: 10.1109/ISCID.2019.00064.
- [27] J. Ramakrishnan, D. Mavaluru, K. Srinivasan, A. Mubarakali, C. Narmatha and G. Malathi, "Opinion Mining using Machine Learning Approaches: A Critical Study," 2020 International Conference on Computing and Information Technology (ICCIT-1441), 2020, pp. 1-4, doi: 10.1109/ICCIT-144147971.2020.9213747.
- [28] G. Li, Q. Zheng, L. Zhang, S. Guo and L. Niu, "Sentiment Infomation based Model For Chinese text Sentiment Analysis," 2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), 2020, pp. 366-371, doi: 10.1109/AUTEEE50969.2020.9315668.
- [29] K. Jia and Z. Li, "Chinese Micro-Blog Sentiment Classification Based on Emotion Dictionary and Semantic Rules," 2020 International Conference on Computer Information and Big Data Applications (CIBDA), 2020, pp. 309-312, doi: 10.1109/CIBDA50819.2020.00076.
- [30] Guezzaz, Y. Asimi, M. Azroul and A. Asimi, "Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection," in Big Data Mining and Analytics, vol. 4, no. 1, pp. 18-24, March 2021, doi: 10.26599/BDMA.2020.9020019.