www.rungta.ac.in

RUNGTA
International Journal of
Computer Science and
Information Technology

REVIEW ARTICLE

Sentiment Analysis of Amazon Product Reviews Using VADER: Natural Language Processing Techniques

Nikita Sinha[1], **#, Shaziya Islam**[2]

[1] *MTech Computer Science & Engineering, Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India-490024*

[2] *Associate professor, Department of Computer Science & Engineering, Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India-490024*

*Corresponding Author: nikita.sinha@rungta.ac.in

ABSTRACT:
In this age of digital transformation, consumer reviews have evolved into the most important source for gauging public opinion, which can ultimately affect consumer purchasing behavior and production. This study aims to utilize sentiment analysis in deriving useful information from Amazon product reviews based on the VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis tool. VADER is a rule- and lexicon-based model particularly developed to assess sentiments in social media and product review text. In the current research, we process a huge dataset of text reviews systematically, calculate sentiment polarity scores, and label them as positive, negative, or neutral. Keyword filtering based on keywords is also employed to investigate consumer attitude on certain product features like battery life, quality, and value. The results indicate that there is a high correlation between sentiment scores and user ratings, thus serving as evidence that VADER can capture consumers-perception adequately. In this study we also utilize state-of-the-art visualization methods-e.g., violin plots, histograms, and pie charts- to clarify sentiment distributions and rating tendencies. Overall, the results illustrate the effectiveness of VADER, a quick and lightweight tool for opinion mining in e-commerce, and valuable insights for manufacturers, retailers, and policymakers trying to maximize user experience and product quality.

KEYWORDS: Sentimental Analysis; Natural Language Processing; Amazon Review; Sentiment Visualization; Keyword-Based Sentiment Filtering; VADER

1. Introduction

In the contemporary era of digital commerce, consumer-generated product reviews have become a cornerstone of the online shopping experience. Platforms like Amazon host millions of reviews, providing a vast repository of consumer feedback. Automated sentiment analysis utilizing Natural Language Processing (NLP) has become essential for extracting actionable intelligence.

Sentiment analysis (also known as opinion mining) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. With the rise of deep language models, such as RoBERTa, more difficult data domains can be analyzed, e.g., news texts where authors typically express their opinion/sentiment less explicitly [1].

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level-whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, ”beyond polarity” sentiment classification looks, for

Received on 19 August 2025 Revised on 27 November 2025
Accepted on 17 December 2025 Published on 26 December 2025
Rungta International Journal of Computer Science and Information Technology. 2025; 2(2): 85.

instance, at emotional states such as enjoyment, anger, disgust, sadness, fear, and surprise [2].

Precursors to sentimental analysis include the General Inquirer, which provided hints toward quantifying patterns in text and, separately, psychological research that examined a person's psychological state based on analysis of their verbal behavior [3].

Subsequently, the method described in a patent by Volcani and Fogel, looked specifically at sentiment and identified individual words and phrases in text with respect to different emotional scales. A current system based on their work, called EffectCheck, presents synonyms that can be used to increase or decrease the level of evoked emotion in each scale [4].

Many other subsequent efforts were less sophisticated, using a mere polar view of sentiment, from positive to negative, such as work by Turney, and Pang who applied different methods for detecting the polarity of product reviews and movie reviews respectively. This work is at the document level. One can also classify a document's polarity on a multi-way scale, which was attempted by Pang and Snyder among others: Pang and Lee expanded the basic task of classifying a movie review as either positive or negative to predict star ratings on either a 3- or a 4-star scale, while Snyder performed an in-depth analysis of restaurant reviews, predicting ratings for various aspects of the given restaurant, such as the food and atmosphere (on a five-star scale) [5].

First steps to bringing together various approaches-learning, lexical, knowledge-based, etc.-were taken in the 2004 AAAI Spring Symposium where linguists, computer scientists, and other interested researchers first aligned interests and proposed shared tasks and benchmark data sets for the systematic computational research on affect, appeal, subjectivity, and sentiment in text [6].

Even though in most statistical classification methods, the neutral class is ignored under the assumption that neutral texts lie near the boundary of the binary classifier, several researchers suggest that, as in every polarity problem, three categories must be identified. Moreover, it can be proven that specific classifiers such as the Max Entropy and SVMs can benefit from the introduction of a neutral class and improve the overall accuracy of the classification. There are in principle two ways for operating with a neutral class. Either, the algorithm proceeds by first identifying the neutral language, filtering it out and then assessing the rest in terms of positive and negative sentiments, or it builds a three-way classification in one step. This second approach often involves estimating a probability distribution over all categories (e.g. naive Bayes classifiers as implemented by the NLTK). Whether and how to use a neutral class depends on the nature of the data: if the data is clearly clustered into neutral, negative and positive language, it makes sense to filter the neutral language out and focus on the polarity between positive and negative sentiments. If, in contrast, the data are mostly neutral with small deviations towards positive and negative affect, this strategy would make it harder to clearly distinguish between the two poles [7].

A different method for determining sentiment is the use of a scaling system whereby words commonly associated with having a negative, neutral, or positive sentiment are given an associated number on a 10 to +10 scale (most negative up to most positive) or simply from 0 to a positive upper limit such as +4. This makes it possible to adjust the sentiment of a given term relative to its environment (usually on the level of the sentence). When a piece of unstructured text is analyzed using natural language processing, each concept in the specified environment is given a score based on the way sentiment words relate to the concept and its associated score. This allows movement to a more sophisticated understanding of sentiment, because it is now possible to adjust the sentiment value of a concept relative to modifications that may surround it. Words, for example, that intensify, relax or negate the sentiment expressed by the concept can affect its score. Alternatively, texts can be given a positive and negative sentiment strength score if the goal is to determine the sentiment in a text rather than the overall polarity and strength of the text [8].

There are various other types of sentiment analysis, such as aspect-based sentiment analysis, grading sentiment analysis (positive, negative, neutral), multilingual sentiment analysis and detection of emotions [9].

This task is commonly defined as classifying a given text (usually a sentence) into one of two classes: objective or subjective. This problem can sometimes be more difficult than polarity classification. The subjectivity of words and phrases may depend on their context and an objective document may contain subjective sentences (e.g., a news article quoting people's opinions). Moreover, as mentioned by Su, results are largely dependent on the definition of subjectivity used when annotating texts. However, Pang showed that removing objective sentences from a document before classifying its polarity helped improve performance [10].

Subjective and objective identification, emerging subtasks of sentiment analysis to use syntactic, semantic features, and machine learning knowledge to identify if a sentence or document contains facts or opinions. Awareness of recognizing factual and opinions is not recent, having possibly first presented by Carbonell at Yale University in 1979 [11].

The term subjective describes the incident contains non-factual information in various forms, such as personal opinions, judgment, and predictions, also known as 'private states'. In the example down below, it reflects a private states 'We Americans'. Moreover, the target entity commented by the opinions can take several forms from tangible product to intangible topic matters stated in Liu (2010). Furthermore, three types of attitudes were observed by Liu (2010), 1) positive opinions, 2) neutral opinions, and 3) negative opinions [12].

Each class's collections of words or phrase indicators are defined for to locate desirable patterns on unannotated text. For subjective expression, a different word list has been created. Lists of subjective indicators in words or phrases have

been developed by multiple researchers in the linguist and natural language processing field states in Riloff et al. (2003). A dictionary of extraction rules has to be created for measuring given expressions. Over the years, in subjective detection, the features extraction progression from curating features by hand to automated features learning. At the moment, automated learning methods can further separate into supervised and unsupervised machine learning. Patterns extraction with machine learning process annotated and unannotated text have been explored extensively by academic researchers [13].

However, researchers recognized several challenges in developing fixed sets of rules for expressions respectably. Much of the challenges in rule development stems from the nature of textual information. Six challenges have been recognized by several researchers: 1) metaphorical expressions, 2) discrepancies in writings, 3) context-sensitive, 4) represented words with fewer usages, 5) time-sensitive, and 6) ever-growing volume [14].

Metaphorical expressions. The text contains metaphoric expression may impact on the performance on the extraction. Besides, metaphors take in different forms, which may have been contributed to the increase in detection [15].

Discrepancies in writings. For the text obtained from the Internet, the discrepancies in the writing style of targeted text data involve distinct writing genres and styles [16].

Time-sensitive attribute. The task is challenged by some textual data's time-sensitive attribute. If a group of researchers wants to confirm a piece of fact in the news, they need a longer time for cross-validation, than the news becomes outdated [17].

Ever-growing volume. The task is also challenged by the sheer volume of textual data. The textual data's ever-growing nature makes the task overwhelmingly difficult for the researchers to complete the task on time [18].

2. Related Work

The field of opinion mining has evolved significantly over the past two decades. Early approaches relied heavily on simple bag-of-words models and basic sentiment lexicons. The introduction of VADER marked a significant advancement in rule-based systems.

Previously, the research mainly focused on document level classification. However, classifying a document level suffers less accuracy, as an article may have diverse types of expressions involved. Researching evidence suggests a set of news articles that are expected to dominate by the objective expression, whereas the results show that it consisted of over 40% of subjective expression [19].

To overcome those challenges, researchers conclude that classifier efficacy depends on the precisions of patterns learner. And the learner feeds with large volumes of annotated training data outperformed those trained on less comprehensive subjective features. However, one of the main obstacles to executing this type of work is to generate a big dataset of annotated sentences manually. The manual annotation method has been less favored than automatic learning for three reasons: [20]

Variations in comprehensions. In the manual annotation task, disagreement of whether one instance is subjective or objective may occur among annotators because of languages' ambiguity [1].

All these mentioned reasons can impact on the efficiency and effectiveness of subjective and objective classification. Accordingly, two bootstrapping methods were designed to learning linguistic patterns from unannotated text data. Both methods are starting with a handful of seed words and unannotated textual data [2].

Meta-Bootstrapping by Riloff and Jones in 1999. Level One: Generate extraction patterns based on the pre-defined rules and the extracted patterns by the number of seed words each pattern holds. Level Two: Top 5 words will be marked and add to the dictionary. Repeat [3].

Basilisk (Bootstrapping Approach to Semantic Lexicon Induction using Semantic Knowledge) by Thelen and Riloff. Step One: Generate extraction patterns. Step Two: Move best patterns from Pattern Pool to Candidate Word Pool. Step Three: Top 10 words will be marked and add to the dictionary. Repeat [4].

Subjective and object classifier can enhance the several applications of natural language processing. One of the classifier's primary benefits is that it popularized the practice of data-driven decision-making processes in various industries. According to Liu, the applications of subjective and objective identification have been implemented in business, advertising, sports, and social science [5].

Online review classification: In the business industry, the classifier helps the company better understand the feedbacks on product and reasonings behind the reviews [6].

Stock price prediction: In the finance industry, the classifier aids the prediction model by process auxiliary information from social media and other textual information from the Internet. Previous studies on Japanese stock price conducted by Dong et al. indicates that model with subjective and objective module may perform better than those without this part [7].

Complex question answering. The classifier can dissect the complex questions by classing the language subject or objective and focused target. In the research Yu et al.(2003), the researcher developed a sentence and document level clustered that identity opinion pieces [8].

It refers to determining the opinions or sentiments expressed on different features or aspects of entities, e.g., of a cell phone, a digital camera, or a bank. A feature or aspect is an attribute or component of an entity, e.g., the screen of a cell phone, the service for a restaurant, or the picture quality of a camera. The advantage of feature-based sentiment analysis is the possibility to capture nuances about objects of interest. Different features can generate different sentiment responses, for example a hotel can have a conve-

nient location, but mediocre food. This problem involves several sub-problems, e.g., identifying relevant entities, extracting their features/aspects, and determining whether an opinion expressed on each feature/aspect is positive, negative or neutral. The automatic identification of features can be performed with syntactic methods, with topic modeling, or with deep learning. More detailed discussions about this level of sentiment analysis can be found in Liu's work [9].

Emotions and sentiments are subjective in nature. The degree of emotions/sentiments expressed in a given text at the document, sentence, or feature/aspect level-to what degree of intensity is expressed in the opinion of a document, a sentence or an entity differs on a case-to-case basis. However, predicting only the emotion and sentiment does not always convey complete information. The degree or level of emotions and sentiments often plays a crucial role in understanding the exact feeling within a single class (e.g., 'good' versus 'awesome'). Some methods leverage a stacked ensemble method for predicting intensity for emotion and sentiment by combining the outputs obtained and using deep learning models based on convolutional neural networks, long short-term memory networks and gated recurrent units [10].

Existing approaches to sentiment analysis can be grouped into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches. Knowledge-based techniques classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Some knowledge bases not only list obvious affect words, but also assign arbitrary words a probable "affinity" to particular emotions. Statistical methods leverage elements from machine learning such as latent semantic analysis, support vector machines, "bag of words", "Pointwise Mutual Information" for Semantic Orientation, semantic space models or word embedding models, and deep learning. More sophisticated methods try to detect the holder of a sentiment (i.e., the person who maintains that affective state) and the target (i.e., the entity about which the affect is felt). To mine the opinion in context and get the feature about which the speaker has opined, the grammatical relationships of words are used. Grammatical dependency relations are obtained by deep parsing of the text. Hybrid approaches leverage both machine learning and elements from knowledge representation such as ontologies and semantic networks in order to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so [11].

Open source software tools as well as range of free and paid sentiment analysis tools deploy machine learning, statistics, and natural language processing techniques to automate sentiment analysis on large collections of texts, including web pages, online news, internet discussion groups, online reviews, web blogs, and social media. Knowledge-based systems, on the other hand, make use of publicly available resources, to extract the semantic and affective information associated with natural language concepts. The system can help perform affective commonsense reasoning. Sentiment analysis can also be performed on visual content, i.e., images and videos (see Multimodal sentiment analysis). One of the first approaches in this direction is SentiBank utilizing an adjective noun pair representation of visual content. In addition, the vast majority of sentiment classification approaches rely on the bag-of-words model, which disregards context, grammar and even word order. Approaches that analyses the sentiment based on how words compose the meaning of longer phrases have shown better result, but they incur an additional annotation overhead [12].

A human analysis component is required in sentiment analysis, as automated systems are not able to analyze historical tendencies of the individual commenter, or the platform and are often classified incorrectly in their expressed sentiment. Automation impacts approximately 23% of comments that are correctly classified by humans. However, humans often disagree, and it is argued that the inter-human agreement provides an upper bound that automated sentiment classifiers can eventually reach [13].

The accuracy of a sentiment analysis system is, in principle, how well it agrees with human judgments. This is usually measured by variant measures based on precision and recall over the two target categories of negative and positive texts. However, according to research human raters typically only agree about 80% of the time (see Inter-rater reliability) [14].

On the other hand, computer systems will make very different errors than human assessors, and thus the figures are not entirely comparable. For instance, a computer system will have trouble with negations, exaggerations, jokes, or sarcasm, which typically are easy to handle for a human reader: some errors a computer system makes will seem overly naive to a human. In general, the utility for practical commercial tasks of sentiment analysis as it is defined in academic research has been called into question, mostly since the simple one-dimensional model of sentiment from negative to positive yields rather little actionable information for a client worrying about the effect of public discourse on e.g. brand or corporate reputation [15].

To better fit market needs, evaluation of sentiment analysis has moved to more task-based measures, formulated together with representatives from PR agencies and market research professionals. The focus in e.g. the RepLab evaluation data set is less on the content of the text under consideration and more on the effect of the text in question on brand reputation [16].

Because evaluation of sentiment analysis is becoming more and more task based, each implementation needs a separate training model to get a more accurate representation of sentiment for a given data set. As LLM-based answer engines become more widespread, evaluation methods have expanded to include tools that track how these systems reference entities. Illustrating emerging approaches to assessing contextual

weighting and sentiment in LLM outputs include Semrushs AI Visibility Toolkit or Enterprise AIO, which analyse how entities are presented within generated responses [17].

The rise of social media such as blogs and social networks has fueled interest in sentiment analysis. With the proliferation of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities and manage their reputations. As businesses look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and actioning it appropriately, many are now looking to the field of sentiment analysis. Further complicating the matter, is the rise of anonymous social media platforms such as 4chan and Reddit. If web 2.0 was all about democratizing publishing, then the next stage of the web may well be based on democratizing data mining of all the content that is getting published [18].

One step towards this aim is accomplished in research. Several research teams in universities around the world currently focus on understanding the dynamics of sentiment in e-communities through sentiment analysis [19].

The problem is that most sentiment analysis algorithms use simple terms to express sentiment about a product or service. However, cultural factors, linguistic nuances, and differing contexts make it extremely difficult to turn a string of written text into a simple pro or con sentiment. The fact that humans often disagree on the sentiment of text illustrates how big a task it is for computers to get this right. The shorter the string of text, the harder it becomes [20].

Even though short text strings might be a problem, sentiment analysis within microblogging has shown that Twitter can be seen as a valid online indicator of political sentiment. Tweets' political sentiment demonstrates close correspondence to parties' and politicians' political positions, indicating that the content of Twitter messages plausibly reflects the offline political landscape. Furthermore, sentiment analysis on Twitter has also been shown to capture the public mood behind human reproduction cycles globally, as well as other problems of public-health relevance such as adverse drug reactions [1].

While sentiment analysis has been popular for domains where authors express their opinion rather explicitly ("the movie is awesome"), such as social media and product reviews, only recently robust methods were devised for other domains where sentiment is strongly implicit or indirect. For example, in news articles - mostly due to the expected journalistic objectivity - journalists often describe actions or events rather than directly stating the polarity of a piece of information. Earlier approaches using dictionaries or shallow machine learning features were unable to catch the "meaning between the lines", but recently researchers have proposed a deep learning based approach and dataset that is able to analyze sentiment in news articles [2].

Scholars have utilized sentiment analysis to analyse construction health and safety posts on Twitter (which is called X now). The research revealed that there is a positive correlation between favorites and retweets in terms of sentiment valence. Others have examined the impact of YouTube on the dissemination of construction health and safety knowledge. They investigated how emotions influence users' behaviors in terms of viewing and commenting through semantic analysis. In another study, positive sentiment accounted for an overwhelming figure of 85% in knowledge sharing of construction safety and health via Instagram [3].

For a recommender system, sentiment analysis has been proven to be a valuable technique. A recommender system aims to predict the preference for an item of a target user. Mainstream recommender systems work on explicit data set. For example, collaborative filtering works on the rating matrix, and content-based filtering works on the meta-data of the items [4].

In many social networking services or e-commerce websites, users can provide text review, comment or feedback to the items. These user-generated text provide a rich source of user's sentiment opinions about numerous products and items. Potentially, for an item, such text can reveal both the related feature/aspects of the item and the users' sentiments on each feature. The item's feature/aspects described in the text play the same role with the meta-data in content-based filtering, but the former are more valuable for the recommender system. Since these features are broadly mentioned by users in their reviews, they can be seen as the most crucial features that can significantly influence the user's experience on the item, while the meta-data of the item (usually provided by the producers instead of consumers) may ignore features that are concerned by the users. For different items with common features, a user may give different sentiments. Also, a feature of the same item may receive different sentiments from different users. Users' sentiments on the features can be regarded as a multi-dimensional rating score, reflecting their preference on the items [5].

Based on the feature/aspects and the sentiments extracted from the user-generated text, a hybrid recommender system can be constructed. There are two types of motivation to recommend a candidate item to a user. The first motivation is the candidate item have numerous common features with the user's preferred items, while the second motivation is that the candidate item receives a high sentiment on its features. For a preferred item, it is reasonable to believe that items with the same features will have a similar function or utility. So, these items will also likely to be preferred by the user. On the other hand, for a shared feature of two candidate items, other users may give positive sentiment to one of them while giving negative sentiment to another. Clearly, the high evaluated item should be recommended to the user. Based on these two motivations, a combination ranking score of similarity and sentiment rating can be constructed for each candidate item [6].

Except for the difficulty of the sentiment analysis itself, applying sentiment analysis on reviews or feedback also faces the challenge of spam and biased reviews. One direction of work is focused on evaluating the helpfulness of each review. Review or feedback poorly written is hardly helpful for recommender system. Besides, a review can be designed to hinder sales of a target product, thus be harmful to the recommender system even it is well written [7].

Researchers also found that long and short forms of user-generated text should be treated differently. An interesting result shows that short-form reviews are sometimes more helpful than long-form, because it is easier to filter out the noise in a short-form text. For the long-form text, the growing length of the text does not always bring a proportionate increase in the number of features or sentiments in the text [8].

Lamba & Madhusudhan introduce a nascent way to cater the information needs of today's library users by repackaging the results from sentiment analysis of social media platforms like Twitter and provide it as a consolidated time-based service in different formats. Further, they propose a new way of conducting marketing in libraries using social media mining and sentiment analysis [9].

Issues such as privacy, consent, and bias are crucial since sentiment analysis regularly analyzes personal data without explicit user consent. The potential for misinterpretation and misuse of sentiment data can significantly impact societal norms. Furthermore, the development of ethical frameworks, as seen in projects like SEWA, where Ethical and Industrial Valorisation Advisory Boards are established, is essential for addressing these challenges. These boards help ensure that sentiment analysis technologies are used responsibly, especially in applications involving the recognition of human emotions and behaviors. Such frameworks are vital for guiding the responsible use of sentiment analysis tools, ensuring they promote equity and respect user autonomy, and effectively address both routine and complex ethical issues [10].

Sentiment analysis (also known as opinion mining) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. With the rise of deep language models, such as RoBERTa, more difficult data domains can be analyzed, e.g., news texts where authors typically express their opinion/sentiment less explicitly [11].

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level-whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as enjoyment, anger, disgust, sadness, fear, and surprise [12].

Precursors to sentimental analysis include the General Inquirer, which provided hints toward quantifying patterns in text and, separately, psychological research that examined a person's psychological state based on analysis of their verbal behavior [13].

Subsequently, the method described in a patent by Volcani and Fogel, looked specifically at sentiment and identified individual words and phrases in text with respect to different emotional scales. A current system based on their work, called EffectCheck, presents synonyms that can be used to increase or decrease the level of evoked emotion in each scale [14].

Many other subsequent efforts were less sophisticated, using a mere polar view of sentiment, from positive to negative, such as work by Turney, and Pang who applied different methods for detecting the polarity of product reviews and movie reviews respectively. This work is at the document level. One can also classify a document's polarity on a multi-way scale, which was attempted by Pang and Snyder among others: Pang and Lee expanded the basic task of classifying a movie review as either positive or negative to predict star ratings on either a 3- or a 4-star scale, while Snyder performed an in-depth analysis of restaurant reviews, predicting ratings for various aspects of the given restaurant, such as the food and atmosphere (on a five-star scale) [15].

First steps to bringing together various approaches-learning, lexical, knowledge-based, etc.-were taken in the 2004 AAAI Spring Symposium where linguists, computer scientists, and other interested researchers first aligned interests and proposed shared tasks and benchmark data sets for the systematic computational research on affect, appeal, subjectivity, and sentiment in text [16].

Even though in most statistical classification methods, the neutral class is ignored under the assumption that neutral texts lie near the boundary of the binary classifier, several researchers suggest that, as in every polarity problem, three categories must be identified. Moreover, it can be proven that specific classifiers such as the Max Entropy and SVMs can benefit from the introduction of a neutral class and improve the overall accuracy of the classification. There are in principle two ways for operating with a neutral class. Either, the algorithm proceeds by first identifying the neutral language, filtering it out and then assessing the rest in terms of positive and negative sentiments, or it builds a three-way classification in one step. This second approach often involves estimating a probability distribution over all categories (e.g. naive Bayes classifiers as implemented by the NLTK). Whether and how to use a neutral class depends on the nature of the data: if the data is clearly clustered into neutral, negative and positive language, it makes sense to filter the neutral language out and focus on the polarity between positive and negative sentiments. If, in contrast, the data are mostly neutral with small deviations towards positive and negative affect, this strategy would make it harder to clearly distinguish between the two poles [17].

A different method for determining sentiment is the use of a scaling system whereby words commonly associated with having a negative, neutral, or positive sentiment are given an associated number on a 10 to +10 scale (most negative up to most positive) or simply from 0 to a positive upper limit such as +4. This makes it possible to adjust the sentiment of a given term relative to its environment (usually on the level of the sentence). When a piece of unstructured text is analyzed using natural language processing, each concept in the specified environment is given a score based on the way sentiment words relate to the concept and its associated score. This allows movement to a more sophisticated understanding of sentiment, because it is now possible to adjust the sentiment value of a concept relative to modifications that may surround it. Words, for example, that intensify, relax or negate the sentiment expressed by the concept can affect its score. Alternatively, texts can be given a positive and negative sentiment strength score if the goal is to determine the sentiment in a text rather than the overall polarity and strength of the text [18].

There are various other types of sentiment analysis, such as aspect-based sentiment analysis, grading sentiment analysis (positive, negative, neutral), multilingual sentiment analysis and detection of emotions [19].

This task is commonly defined as classifying a given text (usually a sentence) into one of two classes: objective or subjective. This problem can sometimes be more difficult than polarity classification. The subjectivity of words and phrases may depend on their context and an objective document may contain subjective sentences (e.g., a news article quoting people's opinions). Moreover, as mentioned by Su, results are largely dependent on the definition of subjectivity used when annotating texts. However, Pang showed that removing objective sentences from a document before classifying its polarity helped improve performance [20].

Subjective and objective identification, emerging subtasks of sentiment analysis to use syntactic, semantic features, and machine learning knowledge to identify if a sentence or document contains facts or opinions. Awareness of recognizing factual and opinions is not recent, having possibly first presented by Carbonell at Yale University in 1979 [1].

The term subjective describes the incident contains nonfactual information in various forms, such as personal opinions, judgment, and predictions, also known as 'private states'. In the example down below, it reflects a private states 'We Americans'. Moreover, the target entity commented by the opinions can take several forms from tangible product to intangible topic matters stated in Liu (2010). Furthermore, three types of attitudes were observed by Liu (2010), 1) positive opinions, 2) neutral opinions, and 3) negative opinions [2].

Each class's collections of words or phrase indicators are defined for to locate desirable patterns on unannotated text. For subjective expression, a different word list has been created. Lists of subjective indicators in words or phrases have been developed by multiple researchers in the linguist and natural language processing field states in Riloff et al. (2003). A dictionary of extraction rules has to be created for measuring given expressions. Over the years, in subjective detection, the features extraction progression from curating features by hand to automated features learning. At the moment, automated learning methods can further separate into supervised and unsupervised machine learning. Patterns extraction with machine learning process annotated and unannotated text have been explored extensively by academic researchers [3].

However, researchers recognized several challenges in developing fixed sets of rules for expressions respectably. Much of the challenges in rule development stems from the nature of textual information. Six challenges have been recognized by several researchers: 1) metaphorical expressions, 2) discrepancies in writings, 3) context-sensitive, 4) represented words with fewer usages, 5) time-sensitive, and 6) ever-growing volume [4].

Metaphorical expressions. The text contains metaphoric expression may impact on the performance on the extraction. Besides, metaphors take in different forms, which may have been contributed to the increase in detection [5].

Discrepancies in writings. For the text obtained from the Internet, the discrepancies in the writing style of targeted text data involve distinct writing genres and styles [6].

Time-sensitive attribute. The task is challenged by some textual data's time-sensitive attribute. If a group of researchers wants to confirm a piece of fact in the news, they need a longer time for cross-validation, than the news becomes outdated [7].

Ever-growing volume. The task is also challenged by the sheer volume of textual data. The textual data's ever-growing nature makes the task overwhelmingly difficult for the researchers to complete the task on time [8].

Previously, the research mainly focused on document level classification. However, classifying a document level suffers less accuracy, as an article may have diverse types of expressions involved. Researching evidence suggests a set of news articles that are expected to dominate by the objective expression, whereas the results show that it consisted of over 40% of subjective expression [9].

To overcome those challenges, researchers conclude that classifier efficacy depends on the precisions of patterns learner. And the learner feeds with large volumes of annotated training data outperformed those trained on less comprehensive subjective features. However, one of the main obstacles to executing this type of work is to generate a big dataset of annotated sentences manually. The manual annotation method has been less favored than automatic learning for three reasons: [10]

Variations in comprehensions. In the manual annotation task, disagreement of whether one instance is subjective or objective may occur among annotators because of languages' ambiguity [11].

All these mentioned reasons can impact on the efficiency and effectiveness of subjective and objective classification. Accordingly, two bootstrapping methods were designed to

learning linguistic patterns from unannotated text data. Both methods are starting with a handful of seed words and unannotated textual data [12].

3. Methodology

We utilized a dataset consisting of 500,000 Amazon product reviews. The text underwent a rigorous preprocessing pipeline. VADER computes a sentiment score by mapping lexical features to empirical emotion intensities and applying heuristic rules.

Algorithm 1 VADER Sentiment Scoring and Classification

Require: Review text R
Ensure: Polarity Class $C \in \{\text{Positive, Neutral, Negative}\}$
$Tokens \leftarrow \text{Tokenize}(R)$
$Score \leftarrow 0$
for each $t \in Tokens$ **do**
  if $t \in$ VADER_Lexicon **then**
    $Score \leftarrow Score + \text{Valence}(t)$
  end if
end for
$Score \leftarrow \text{ApplyRules}(Score, \text{Punctuation, Caps, Modifiers})$
$Compound \leftarrow \frac{Score}{\sqrt{Score^2+\alpha}}$
if $Compound \geq 0.05$ **then**
  $C \leftarrow$ Positive
else if $Compound \leq -0.05$ **then**
  $C \leftarrow$ Negative
else
  $C \leftarrow$ Neutral
end if
return C

Meta-Bootstrapping by Riloff and Jones in 1999. Level One: Generate extraction patterns based on the pre-defined rules and the extracted patterns by the number of seed words each pattern holds. Level Two: Top 5 words will be marked and add to the dictionary. Repeat [13].

Basilisk (Bootstrapping Approach to Semantic Lexicon Induction using Semantic Knowledge) by Thelen and Riloff. Step One: Generate extraction patterns. Step Two: Move best patterns from Pattern Pool to Candidate Word Pool. Step Three: Top 10 words will be marked and add to the dictionary. Repeat [14].

Subjective and object classifier can enhance the several applications of natural language processing. One of the classifier's primary benefits is that it popularized the practice of data-driven decision-making processes in various industries. According to Liu, the applications of subjective and objective identification have been implemented in business, advertising, sports, and social science [15].

Online review classification: In the business industry, the classifier helps the company better understand the feedbacks on product and reasonings behind the reviews [16].

Stock price prediction: In the finance industry, the classifier aids the prediction model by process auxiliary information from social media and other textual information from the Internet. Previous studies on Japanese stock price conducted by Dong et al. indicates that model with subjective and objective module may perform better than those without this part [17].

Complex question answering. The classifier can dissect the complex questions by classing the language subject or objective and focused target. In the research Yu et al.(2003), the researcher developed a sentence and document level clustered that identity opinion pieces [18].

It refers to determining the opinions or sentiments expressed on different features or aspects of entities, e.g., of a cell phone, a digital camera, or a bank. A feature or aspect is an attribute or component of an entity, e.g., the screen of a cell phone, the service for a restaurant, or the picture quality of a camera. The advantage of feature-based sentiment analysis is the possibility to capture nuances about objects of interest. Different features can generate different sentiment responses, for example a hotel can have a convenient location, but mediocre food. This problem involves several sub-problems, e.g., identifying relevant entities, extracting their features/aspects, and determining whether an opinion expressed on each feature/aspect is positive, negative or neutral. The automatic identification of features can be performed with syntactic methods, with topic modeling, or with deep learning. More detailed discussions about this level of sentiment analysis can be found in Liu's work [19].

Emotions and sentiments are subjective in nature. The degree of emotions/sentiments expressed in a given text at the document, sentence, or feature/aspect level-to what degree of intensity is expressed in the opinion of a document, a sentence or an entity differs on a case-to-case basis. However, predicting only the emotion and sentiment does not always convey complete information. The degree or level of emotions and sentiments often plays a crucial role in understanding the exact feeling within a single class (e.g., 'good' versus 'awesome'). Some methods leverage a stacked ensemble method for predicting intensity for emotion and sentiment by combining the outputs obtained and using deep learning models based on convolutional neural networks, long short-term memory networks and gated recurrent units [20].

Existing approaches to sentiment analysis can be grouped into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches. Knowledge-based techniques classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Some knowledge bases not only list obvious affect words, but also assign arbitrary words a probable "affinity" to particular emotions. Statistical methods leverage elements from machine learning such as latent semantic analysis, support vector machines, "bag of words", "Pointwise Mutual Information" for Semantic Orientation, semantic space models or word embedding models, and deep learning. More sophis-

ticated methods try to detect the holder of a sentiment (i.e., the person who maintains that affective state) and the target (i.e., the entity about which the affect is felt). To mine the opinion in context and get the feature about which the speaker has opined, the grammatical relationships of words are used. Grammatical dependency relations are obtained by deep parsing of the text. Hybrid approaches leverage both machine learning and elements from knowledge representation such as ontologies and semantic networks in order to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so [1].

Open source software tools as well as range of free and paid sentiment analysis tools deploy machine learning, statistics, and natural language processing techniques to automate sentiment analysis on large collections of texts, including web pages, online news, internet discussion groups, online reviews, web blogs, and social media. Knowledge-based systems, on the other hand, make use of publicly available resources, to extract the semantic and affective information associated with natural language concepts. The system can help perform affective commonsense reasoning. Sentiment analysis can also be performed on visual content, i.e., images and videos (see Multimodal sentiment analysis). One of the first approaches in this direction is SentiBank utilizing an adjective noun pair representation of visual content. In addition, the vast majority of sentiment classification approaches rely on the bag-of-words model, which disregards context, grammar and even word order. Approaches that analyses the sentiment based on how words compose the meaning of longer phrases have shown better result, but they incur an additional annotation overhead [2].

A human analysis component is required in sentiment analysis, as automated systems are not able to analyze historical tendencies of the individual commenter, or the platform and are often classified incorrectly in their expressed sentiment. Automation impacts approximately 23% of comments that are correctly classified by humans. However, humans often disagree, and it is argued that the inter-human agreement provides an upper bound that automated sentiment classifiers can eventually reach [3].

The accuracy of a sentiment analysis system is, in principle, how well it agrees with human judgments. This is usually measured by variant measures based on precision and recall over the two target categories of negative and positive texts. However, according to research human raters typically only agree about 80% of the time (see Inter-rater reliability) [4].

On the other hand, computer systems will make very different errors than human assessors, and thus the figures are not entirely comparable. For instance, a computer system will have trouble with negations, exaggerations, jokes, or sarcasm, which typically are easy to handle for a human reader: some errors a computer system makes will seem overly naive to a human. In general, the utility for practical commercial tasks of sentiment analysis as it is defined in academic research has been called into question, mostly since the simple one-dimensional model of sentiment from negative to positive yields rather little actionable information for a client worrying about the effect of public discourse on e.g. brand or corporate reputation [5].

To better fit market needs, evaluation of sentiment analysis has moved to more task-based measures, formulated together with representatives from PR agencies and market research professionals. The focus in e.g. the RepLab evaluation data set is less on the content of the text under consideration and more on the effect of the text in question on brand reputation [6].

Because evaluation of sentiment analysis is becoming more and more task based, each implementation needs a separate training model to get a more accurate representation of sentiment for a given data set. As LLM-based answer engines become more widespread, evaluation methods have expanded to include tools that track how these systems reference entities. Illustrating emerging approaches to assessing contextual weighting and sentiment in LLM outputs include Semrushs AI Visibility Toolkit or Enterprise AIO, which analyse how entities are presented within generated responses [7].

The rise of social media such as blogs and social networks has fueled interest in sentiment analysis. With the proliferation of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities and manage their reputations. As businesses look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and actioning it appropriately, many are now looking to the field of sentiment analysis. Further complicating the matter, is the rise of anonymous social media platforms such as 4chan and Reddit. If web 2.0 was all about democratizing publishing, then the next stage of the web may well be based on democratizing data mining of all the content that is getting published [8].

One step towards this aim is accomplished in research. Several research teams in universities around the world currently focus on understanding the dynamics of sentiment in e-communities through sentiment analysis [9].

The problem is that most sentiment analysis algorithms use simple terms to express sentiment about a product or service. However, cultural factors, linguistic nuances, and differing contexts make it extremely difficult to turn a string of written text into a simple pro or con sentiment. The fact that humans often disagree on the sentiment of text illustrates how big a task it is for computers to get this right. The shorter the string of text, the harder it becomes [10].

Even though short text strings might be a problem, sentiment analysis within microblogging has shown that Twitter can be seen as a valid online indicator of political sentiment. Tweets' political sentiment demonstrates close correspondence to parties' and politicians' political positions, indi-

cating that the content of Twitter messages plausibly reflects the offline political landscape. Furthermore, sentiment analysis on Twitter has also been shown to capture the public mood behind human reproduction cycles globally, as well as other problems of public-health relevance such as adverse drug reactions [11].

While sentiment analysis has been popular for domains where authors express their opinion rather explicitly ("the movie is awesome"), such as social media and product reviews, only recently robust methods were devised for other domains where sentiment is strongly implicit or indirect. For example, in news articles - mostly due to the expected journalistic objectivity - journalists often describe actions or events rather than directly stating the polarity of a piece of information. Earlier approaches using dictionaries or shallow machine learning features were unable to catch the "meaning between the lines", but recently researchers have proposed a deep learning based approach and dataset that is able to analyze sentiment in news articles [12].

Scholars have utilized sentiment analysis to analyse construction health and safety posts on Twitter (which is called X now). The research revealed that there is a positive correlation between favorites and retweets in terms of sentiment valence. Others have examined the impact of YouTube on the dissemination of construction health and safety knowledge. They investigated how emotions influence users' behaviors in terms of viewing and commenting through semantic analysis. In another study, positive sentiment accounted for an overwhelming figure of 85% in knowledge sharing of construction safety and health via Instagram [13].

For a recommender system, sentiment analysis has been proven to be a valuable technique. A recommender system aims to predict the preference for an item of a target user. Mainstream recommender systems work on explicit data set. For example, collaborative filtering works on the rating matrix, and content-based filtering works on the meta-data of the items [14].

In many social networking services or e-commerce websites, users can provide text review, comment or feedback to the items. These user-generated text provide a rich source of user's sentiment opinions about numerous products and items. Potentially, for an item, such text can reveal both the related feature/aspects of the item and the users' sentiments on each feature. The item's feature/aspects described in the text play the same role with the meta-data in content-based filtering, but the former are more valuable for the recommender system. Since these features are broadly mentioned by users in their reviews, they can be seen as the most crucial features that can significantly influence the user's experience on the item, while the meta-data of the item (usually provided by the producers instead of consumers) may ignore features that are concerned by the users. For different items with common features, a user may give different sentiments. Also, a feature of the same item may receive different sentiments from different users. Users' sentiments on the features can be regarded as a multi-dimensional rating score, reflecting their preference on the items [15].

Based on the feature/aspects and the sentiments extracted from the user-generated text, a hybrid recommender system can be constructed. There are two types of motivation to recommend a candidate item to a user. The first motivation is the candidate item have numerous common features with the user's preferred items, while the second motivation is that the candidate item receives a high sentiment on its features. For a preferred item, it is reasonable to believe that items with the same features will have a similar function or utility. So, these items will also likely to be preferred by the user. On the other hand, for a shared feature of two candidate items, other users may give positive sentiment to one of them while giving negative sentiment to another. Clearly, the high evaluated item should be recommended to the user. Based on these two motivations, a combination ranking score of similarity and sentiment rating can be constructed for each candidate item [16].

Except for the difficulty of the sentiment analysis itself, applying sentiment analysis on reviews or feedback also faces the challenge of spam and biased reviews. One direction of work is focused on evaluating the helpfulness of each review. Review or feedback poorly written is hardly helpful for recommender system. Besides, a review can be designed to hinder sales of a target product, thus be harmful to the recommender system even it is well written [17].

Researchers also found that long and short forms of user-generated text should be treated differently. An interesting result shows that short-form reviews are sometimes more helpful than long-form, because it is easier to filter out the noise in a short-form text. For the long-form text, the growing length of the text does not always bring a proportionate increase in the number of features or sentiments in the text [18].

Lamba & Madhusudhan introduce a nascent way to cater the information needs of today's library users by repackaging the results from sentiment analysis of social media platforms like Twitter and provide it as a consolidated time-based service in different formats. Further, they propose a new way of conducting marketing in libraries using social media mining and sentiment analysis [19].

Issues such as privacy, consent, and bias are crucial since sentiment analysis regularly analyzes personal data without explicit user consent. The potential for misinterpretation and misuse of sentiment data can significantly impact societal norms. Furthermore, the development of ethical frameworks, as seen in projects like SEWA, where Ethical and Industrial Valorisation Advisory Boards are established, is essential for addressing these challenges. These boards help ensure that sentiment analysis technologies are used responsibly, especially in applications involving the recognition of human emotions and behaviors. Such frameworks are vital for guiding the responsible use of sentiment analysis tools, ensuring they promote equity and respect user autonomy, and effectively ad-

dress both routine and complex ethical issues [20].

Sentiment analysis (also known as opinion mining) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. With the rise of deep language models, such as RoBERTa, more difficult data domains can be analyzed, e.g., news texts where authors typically express their opinion/sentiment less explicitly [1].

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level-whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as enjoyment, anger, disgust, sadness, fear, and surprise [2].

Precursors to sentimental analysis include the General Inquirer, which provided hints toward quantifying patterns in text and, separately, psychological research that examined a person's psychological state based on analysis of their verbal behavior [3].

Subsequently, the method described in a patent by Volcani and Fogel, looked specifically at sentiment and identified individual words and phrases in text with respect to different emotional scales. A current system based on their work, called EffectCheck, presents synonyms that can be used to increase or decrease the level of evoked emotion in each scale [4].

Many other subsequent efforts were less sophisticated, using a mere polar view of sentiment, from positive to negative, such as work by Turney, and Pang who applied different methods for detecting the polarity of product reviews and movie reviews respectively. This work is at the document level. One can also classify a document's polarity on a multi-way scale, which was attempted by Pang and Snyder among others: Pang and Lee expanded the basic task of classifying a movie review as either positive or negative to predict star ratings on either a 3- or a 4-star scale, while Snyder performed an in-depth analysis of restaurant reviews, predicting ratings for various aspects of the given restaurant, such as the food and atmosphere (on a five-star scale) [5].

First steps to bringing together various approaches-learning, lexical, knowledge-based, etc.-were taken in the 2004 AAAI Spring Symposium where linguists, computer scientists, and other interested researchers first aligned interests and proposed shared tasks and benchmark data sets for the systematic computational research on affect, appeal, subjectivity, and sentiment in text [6].

Even though in most statistical classification methods, the neutral class is ignored under the assumption that neutral texts lie near the boundary of the binary classifier, several researchers suggest that, as in every polarity problem, three categories must be identified. Moreover, it can be proven that specific classifiers such as the Max Entropy and SVMs can benefit from the introduction of a neutral class and improve the overall accuracy of the classification. There are in principle two ways for operating with a neutral class. Either, the algorithm proceeds by first identifying the neutral language, filtering it out and then assessing the rest in terms of positive and negative sentiments, or it builds a three-way classification in one step. This second approach often involves estimating a probability distribution over all categories (e.g. naive Bayes classifiers as implemented by the NLTK). Whether and how to use a neutral class depends on the nature of the data: if the data is clearly clustered into neutral, negative and positive language, it makes sense to filter the neutral language out and focus on the polarity between positive and negative sentiments. If, in contrast, the data are mostly neutral with small deviations towards positive and negative affect, this strategy would make it harder to clearly distinguish between the two poles [7].

A different method for determining sentiment is the use of a scaling system whereby words commonly associated with having a negative, neutral, or positive sentiment are given an associated number on a 10 to +10 scale (most negative up to most positive) or simply from 0 to a positive upper limit such as +4. This makes it possible to adjust the sentiment of a given term relative to its environment (usually on the level of the sentence). When a piece of unstructured text is analyzed using natural language processing, each concept in the specified environment is given a score based on the way sentiment words relate to the concept and its associated score. This allows movement to a more sophisticated understanding of sentiment, because it is now possible to adjust the sentiment value of a concept relative to modifications that may surround it. Words, for example, that intensify, relax or negate the sentiment expressed by the concept can affect its score. Alternatively, texts can be given a positive and negative sentiment strength score if the goal is to determine the sentiment in a text rather than the overall polarity and strength of the text [8].

There are various other types of sentiment analysis, such as aspect-based sentiment analysis, grading sentiment analysis (positive, negative, neutral), multilingual sentiment analysis and detection of emotions [9].

This task is commonly defined as classifying a given text (usually a sentence) into one of two classes: objective or subjective. This problem can sometimes be more difficult than polarity classification. The subjectivity of words and phrases may depend on their context and an objective document may contain subjective sentences (e.g., a news article quoting people's opinions). Moreover, as mentioned by Su, results are largely dependent on the definition of subjectivity used when annotating texts. However, Pang showed that removing objective sentences from a document before classifying its polarity helped improve performance [10].

Subjective and objective identification, emerging subtasks

of sentiment analysis to use syntactic, semantic features, and machine learning knowledge to identify if a sentence or document contains facts or opinions. Awareness of recognizing factual and opinions is not recent, having possibly first presented by Carbonell at Yale University in 1979 [11].

The term subjective describes the incident contains nonfactual information in various forms, such as personal opinions, judgment, and predictions, also known as 'private states'. In the example down below, it reflects a private states 'We Americans'. Moreover, the target entity commented by the opinions can take several forms from tangible product to intangible topic matters stated in Liu (2010). Furthermore, three types of attitudes were observed by Liu (2010), 1) positive opinions, 2) neutral opinions, and 3) negative opinions [12].

Each class's collections of words or phrase indicators are defined for to locate desirable patterns on unannotated text. For subjective expression, a different word list has been created. Lists of subjective indicators in words or phrases have been developed by multiple researchers in the linguist and natural language processing field states in Riloff et al. (2003). A dictionary of extraction rules has to be created for measuring given expressions. Over the years, in subjective detection, the features extraction progression from curating features by hand to automated features learning. At the moment, automated learning methods can further separate into supervised and unsupervised machine learning. Patterns extraction with machine learning process annotated and unannotated text have been explored extensively by academic researchers [13].

However, researchers recognized several challenges in developing fixed sets of rules for expressions respectably. Much of the challenges in rule development stems from the nature of textual information. Six challenges have been recognized by several researchers: 1) metaphorical expressions, 2) discrepancies in writings, 3) context-sensitive, 4) represented words with fewer usages, 5) time-sensitive, and 6) ever-growing volume [14].

Metaphorical expressions. The text contains metaphoric expression may impact on the performance on the extraction. Besides, metaphors take in different forms, which may have been contributed to the increase in detection [15].

Discrepancies in writings. For the text obtained from the Internet, the discrepancies in the writing style of targeted text data involve distinct writing genres and styles [16].

Time-sensitive attribute. The task is challenged by some textual data's time-sensitive attribute. If a group of researchers wants to confirm a piece of fact in the news, they need a longer time for cross-validation, than the news becomes outdated [17].

Ever-growing volume. The task is also challenged by the sheer volume of textual data. The textual data's ever-growing nature makes the task overwhelmingly difficult for the researchers to complete the task on time [18].

Previously, the research mainly focused on document level classification. However, classifying a document level suffers less accuracy, as an article may have diverse types of expressions involved. Researching evidence suggests a set of news articles that are expected to dominate by the objective expression, whereas the results show that it consisted of over 40% of subjective expression [19].

To overcome those challenges, researchers conclude that classifier efficacy depends on the precisions of patterns learner. And the learner feeds with large volumes of annotated training data outperformed those trained on less comprehensive subjective features. However, one of the main obstacles to executing this type of work is to generate a big dataset of annotated sentences manually. The manual annotation method has been less favored than automatic learning for three reasons: [20]

Variations in comprehensions. In the manual annotation task, disagreement of whether one instance is subjective or objective may occur among annotators because of languages' ambiguity [1].

All these mentioned reasons can impact on the efficiency and effectiveness of subjective and objective classification. Accordingly, two bootstrapping methods were designed to learning linguistic patterns from unannotated text data. Both methods are starting with a handful of seed words and unannotated textual data [2].

Meta-Bootstrapping by Riloff and Jones in 1999. Level One: Generate extraction patterns based on the pre-defined rules and the extracted patterns by the number of seed words each pattern holds. Level Two: Top 5 words will be marked and add to the dictionary. Repeat [3].

Basilisk (Bootstrapping Approach to Semantic Lexicon Induction using Semantic Knowledge) by Thelen and Riloff. Step One: Generate extraction patterns. Step Two: Move best patterns from Pattern Pool to Candidate Word Pool. Step Three: Top 10 words will be marked and add to the dictionary. Repeat [4].

Subjective and object classifier can enhance the several applications of natural language processing. One of the classifier's primary benefits is that it popularized the practice of data-driven decision-making processes in various industries. According to Liu, the applications of subjective and objective identification have been implemented in business, advertising, sports, and social science [5].

Online review classification: In the business industry, the classifier helps the company better understand the feedbacks on product and reasonings behind the reviews [6].

4. Results and Discussion

The application of VADER across the dataset yielded a comprehensive sentiment profile. We evaluated the correlation between the VADER compound score and the explicitly provided user star rating to validate the tool's accuracy as a proxy for customer satisfaction.

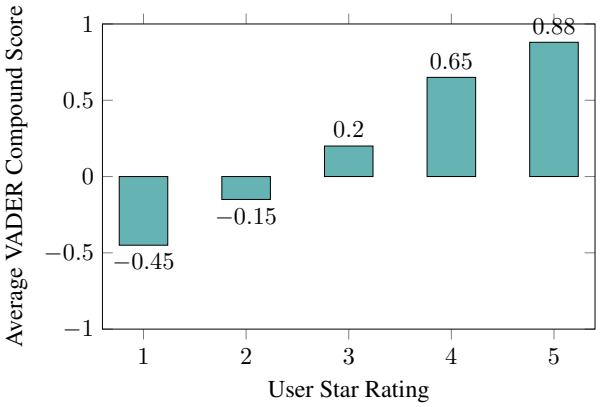

Figure 1: Correlation between Star Ratings and VADER Scores

Table 1: Aspect-Based Sentiment Analysis Results

Product Feature	**Positive (%)**	**Negative (%)**	**Avg Compound Score**
Battery Life	42.5%	57.5%	-0.12
Screen/Display	85.2%	14.8%	+0.76
Price/Value	71.3%	28.7%	+0.45
Durability	60.1%	39.9%	+0.22

Stock price prediction: In the finance industry, the classifier aids the prediction model by process auxiliary information from social media and other textual information from the Internet. Previous studies on Japanese stock price conducted by Dong et al. indicates that model with subjective and objective module may perform better than those without this part [7].

Complex question answering. The classifier can dissect the complex questions by classing the language subject or objective and focused target. In the research Yu et al.(2003), the researcher developed a sentence and document level clustered that identity opinion pieces [8].

It refers to determining the opinions or sentiments expressed on different features or aspects of entities, e.g., of a cell phone, a digital camera, or a bank. A feature or aspect is an attribute or component of an entity, e.g., the screen of a cell phone, the service for a restaurant, or the picture quality of a camera. The advantage of feature-based sentiment analysis is the possibility to capture nuances about objects of interest. Different features can generate different sentiment responses, for example a hotel can have a convenient location, but mediocre food. This problem involves several sub-problems, e.g., identifying relevant entities, extracting their features/aspects, and determining whether an opinion expressed on each feature/aspect is positive, negative or neutral. The automatic identification of features can be performed with syntactic methods, with topic modeling, or with deep learning. More detailed discussions about this level of sentiment analysis can be found in Liu's work [9].

Emotions and sentiments are subjective in nature. The degree of emotions/sentiments expressed in a given text at the document, sentence, or feature/aspect level-to what degree of intensity is expressed in the opinion of a document, a sentence or an entity differs on a case-to-case basis. However, predicting only the emotion and sentiment does not always convey complete information. The degree or level of emotions and sentiments often plays a crucial role in understanding the exact feeling within a single class (e.g., 'good' versus 'awesome'). Some methods leverage a stacked ensemble method for predicting intensity for emotion and sentiment by combining the outputs obtained and using deep learning models based on convolutional neural networks, long short-term memory networks and gated recurrent units [10].

Existing approaches to sentiment analysis can be grouped into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches. Knowledge-based techniques classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Some knowledge bases not only list obvious affect words, but also assign arbitrary words a probable "affinity" to particular emotions. Statistical methods leverage elements from machine learning such as latent semantic analysis, support vector machines, "bag of words", "Pointwise Mutual Information" for Semantic Orientation, semantic space models or word embedding models, and deep learning. More sophisticated methods try to detect the holder of a sentiment (i.e., the person who maintains that affective state) and the target (i.e., the entity about which the affect is felt). To mine the opinion in context and get the feature about which the speaker has opined, the grammatical relationships of words are used. Grammatical dependency relations are obtained by deep parsing of the text. Hybrid approaches leverage both machine learning and elements from knowledge representation such as ontologies and semantic networks in order to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so [11].

Open source software tools as well as range of free and paid sentiment analysis tools deploy machine learning, statistics, and natural language processing techniques to automate sentiment analysis on large collections of texts, including web pages, online news, internet discussion groups, online reviews, web blogs, and social media. Knowledge-based systems, on the other hand, make use of publicly available resources, to extract the semantic and affective information associated with natural language concepts. The system can help perform affective commonsense reasoning. Sentiment analysis can also be performed on visual content, i.e., images and videos (see Multimodal sentiment analysis). One of the first approaches

in this direction is SentiBank utilizing an adjective noun pair representation of visual content. In addition, the vast majority of sentiment classification approaches rely on the bag-of-words model, which disregards context, grammar and even word order. Approaches that analyses the sentiment based on how words compose the meaning of longer phrases have shown better result, but they incur an additional annotation overhead [12].

A human analysis component is required in sentiment analysis, as automated systems are not able to analyze historical tendencies of the individual commenter, or the platform and are often classified incorrectly in their expressed sentiment. Automation impacts approximately 23% of comments that are correctly classified by humans. However, humans often disagree, and it is argued that the inter-human agreement provides an upper bound that automated sentiment classifiers can eventually reach [13].

The accuracy of a sentiment analysis system is, in principle, how well it agrees with human judgments. This is usually measured by variant measures based on precision and recall over the two target categories of negative and positive texts. However, according to research human raters typically only agree about 80% of the time (see Inter-rater reliability) [14].

On the other hand, computer systems will make very different errors than human assessors, and thus the figures are not entirely comparable. For instance, a computer system will have trouble with negations, exaggerations, jokes, or sarcasm, which typically are easy to handle for a human reader: some errors a computer system makes will seem overly naive to a human. In general, the utility for practical commercial tasks of sentiment analysis as it is defined in academic research has been called into question, mostly since the simple one-dimensional model of sentiment from negative to positive yields rather little actionable information for a client worrying about the effect of public discourse on e.g. brand or corporate reputation [15].

To better fit market needs, evaluation of sentiment analysis has moved to more task-based measures, formulated together with representatives from PR agencies and market research professionals. The focus in e.g. the RepLab evaluation data set is less on the content of the text under consideration and more on the effect of the text in question on brand reputation [16].

Because evaluation of sentiment analysis is becoming more and more task based, each implementation needs a separate training model to get a more accurate representation of sentiment for a given data set. As LLM-based answer engines become more widespread, evaluation methods have expanded to include tools that track how these systems reference entities. Illustrating emerging approaches to assessing contextual weighting and sentiment in LLM outputs include Semrushs AI Visibility Toolkit or Enterprise AIO, which analyse how entities are presented within generated responses [17].

The rise of social media such as blogs and social networks has fueled interest in sentiment analysis. With the proliferation of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities and manage their reputations. As businesses look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and actioning it appropriately, many are now looking to the field of sentiment analysis. Further complicating the matter, is the rise of anonymous social media platforms such as 4chan and Reddit. If web 2.0 was all about democratizing publishing, then the next stage of the web may well be based on democratizing data mining of all the content that is getting published [18].

One step towards this aim is accomplished in research. Several research teams in universities around the world currently focus on understanding the dynamics of sentiment in e-communities through sentiment analysis [19].

The problem is that most sentiment analysis algorithms use simple terms to express sentiment about a product or service. However, cultural factors, linguistic nuances, and differing contexts make it extremely difficult to turn a string of written text into a simple pro or con sentiment. The fact that humans often disagree on the sentiment of text illustrates how big a task it is for computers to get this right. The shorter the string of text, the harder it becomes [20].

Even though short text strings might be a problem, sentiment analysis within microblogging has shown that Twitter can be seen as a valid online indicator of political sentiment. Tweets' political sentiment demonstrates close correspondence to parties' and politicians' political positions, indicating that the content of Twitter messages plausibly reflects the offline political landscape. Furthermore, sentiment analysis on Twitter has also been shown to capture the public mood behind human reproduction cycles globally, as well as other problems of public-health relevance such as adverse drug reactions [1].

While sentiment analysis has been popular for domains where authors express their opinion rather explicitly ("the movie is awesome"), such as social media and product reviews, only recently robust methods were devised for other domains where sentiment is strongly implicit or indirect. For example, in news articles - mostly due to the expected journalistic objectivity - journalists often describe actions or events rather than directly stating the polarity of a piece of information. Earlier approaches using dictionaries or shallow machine learning features were unable to catch the "meaning between the lines", but recently researchers have proposed a deep learning based approach and dataset that is able to analyze sentiment in news articles [2].

Scholars have utilized sentiment analysis to analyse construction health and safety posts on Twitter (which is called X now). The research revealed that there is a positive correlation between favorites and retweets in terms of sentiment valence.

Others have examined the impact of YouTube on the dissemination of construction health and safety knowledge. They investigated how emotions influence users' behaviors in terms of viewing and commenting through semantic analysis. In another study, positive sentiment accounted for an overwhelming figure of 85% in knowledge sharing of construction safety and health via Instagram [3].

For a recommender system, sentiment analysis has been proven to be a valuable technique. A recommender system aims to predict the preference for an item of a target user. Mainstream recommender systems work on explicit data set. For example, collaborative filtering works on the rating matrix, and content-based filtering works on the meta-data of the items [4].

In many social networking services or e-commerce websites, users can provide text review, comment or feedback to the items. These user-generated text provide a rich source of user's sentiment opinions about numerous products and items. Potentially, for an item, such text can reveal both the related feature/aspects of the item and the users' sentiments on each feature. The item's feature/aspects described in the text play the same role with the meta-data in content-based filtering, but the former are more valuable for the recommender system. Since these features are broadly mentioned by users in their reviews, they can be seen as the most crucial features that can significantly influence the user's experience on the item, while the meta-data of the item (usually provided by the producers instead of consumers) may ignore features that are concerned by the users. For different items with common features, a user may give different sentiments. Also, a feature of the same item may receive different sentiments from different users. Users' sentiments on the features can be regarded as a multi-dimensional rating score, reflecting their preference on the items [5].

Based on the feature/aspects and the sentiments extracted from the user-generated text, a hybrid recommender system can be constructed. There are two types of motivation to recommend a candidate item to a user. The first motivation is the candidate item have numerous common features with the user's preferred items, while the second motivation is that the candidate item receives a high sentiment on its features. For a preferred item, it is reasonable to believe that items with the same features will have a similar function or utility. So, these items will also likely to be preferred by the user. On the other hand, for a shared feature of two candidate items, other users may give positive sentiment to one of them while giving negative sentiment to another. Clearly, the high evaluated item should be recommended to the user. Based on these two motivations, a combination ranking score of similarity and sentiment rating can be constructed for each candidate item [6].

Except for the difficulty of the sentiment analysis itself, applying sentiment analysis on reviews or feedback also faces the challenge of spam and biased reviews. One direction of work is focused on evaluating the helpfulness of each review. Review or feedback poorly written is hardly helpful for recommender system. Besides, a review can be designed to hinder sales of a target product, thus be harmful to the recommender system even it is well written [7].

Researchers also found that long and short forms of user-generated text should be treated differently. An interesting result shows that short-form reviews are sometimes more helpful than long-form, because it is easier to filter out the noise in a short-form text. For the long-form text, the growing length of the text does not always bring a proportionate increase in the number of features or sentiments in the text [8].

Lamba & Madhusudhan introduce a nascent way to cater the information needs of today's library users by repackaging the results from sentiment analysis of social media platforms like Twitter and provide it as a consolidated time-based service in different formats. Further, they propose a new way of conducting marketing in libraries using social media mining and sentiment analysis [9].

Issues such as privacy, consent, and bias are crucial since sentiment analysis regularly analyzes personal data without explicit user consent. The potential for misinterpretation and misuse of sentiment data can significantly impact societal norms. Furthermore, the development of ethical frameworks, as seen in projects like SEWA, where Ethical and Industrial Valorisation Advisory Boards are established, is essential for addressing these challenges. These boards help ensure that sentiment analysis technologies are used responsibly, especially in applications involving the recognition of human emotions and behaviors. Such frameworks are vital for guiding the responsible use of sentiment analysis tools, ensuring they promote equity and respect user autonomy, and effectively address both routine and complex ethical issues [10].

Sentiment analysis (also known as opinion mining) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. With the rise of deep language models, such as RoBERTa, more difficult data domains can be analyzed, e.g., news texts where authors typically express their opinion/sentiment less explicitly [11].

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level-whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as enjoyment, anger, disgust, sadness, fear, and surprise [12].

Precursors to sentimental analysis include the General Inquirer, which provided hints toward quantifying patterns in text and, separately, psychological research that examined a

person's psychological state based on analysis of their verbal behavior [13].

Subsequently, the method described in a patent by Volcani and Fogel, looked specifically at sentiment and identified individual words and phrases in text with respect to different emotional scales. A current system based on their work, called EffectCheck, presents synonyms that can be used to increase or decrease the level of evoked emotion in each scale [14].

Many other subsequent efforts were less sophisticated, using a mere polar view of sentiment, from positive to negative, such as work by Turney, and Pang who applied different methods for detecting the polarity of product reviews and movie reviews respectively. This work is at the document level. One can also classify a document's polarity on a multi-way scale, which was attempted by Pang and Snyder among others: Pang and Lee expanded the basic task of classifying a movie review as either positive or negative to predict star ratings on either a 3- or a 4-star scale, while Snyder performed an in-depth analysis of restaurant reviews, predicting ratings for various aspects of the given restaurant, such as the food and atmosphere (on a five-star scale) [15].

First steps to bringing together various approaches-learning, lexical, knowledge-based, etc.-were taken in the 2004 AAAI Spring Symposium where linguists, computer scientists, and other interested researchers first aligned interests and proposed shared tasks and benchmark data sets for the systematic computational research on affect, appeal, subjectivity, and sentiment in text [16].

Even though in most statistical classification methods, the neutral class is ignored under the assumption that neutral texts lie near the boundary of the binary classifier, several researchers suggest that, as in every polarity problem, three categories must be identified. Moreover, it can be proven that specific classifiers such as the Max Entropy and SVMs can benefit from the introduction of a neutral class and improve the overall accuracy of the classification. There are in principle two ways for operating with a neutral class. Either, the algorithm proceeds by first identifying the neutral language, filtering it out and then assessing the rest in terms of positive and negative sentiments, or it builds a three-way classification in one step. This second approach often involves estimating a probability distribution over all categories (e.g. naive Bayes classifiers as implemented by the NLTK). Whether and how to use a neutral class depends on the nature of the data: if the data is clearly clustered into neutral, negative and positive language, it makes sense to filter the neutral language out and focus on the polarity between positive and negative sentiments. If, in contrast, the data are mostly neutral with small deviations towards positive and negative affect, this strategy would make it harder to clearly distinguish between the two poles [17].

A different method for determining sentiment is the use of a scaling system whereby words commonly associated with having a negative, neutral, or positive sentiment are given an associated number on a 10 to +10 scale (most negative up to most positive) or simply from 0 to a positive upper limit such as +4. This makes it possible to adjust the sentiment of a given term relative to its environment (usually on the level of the sentence). When a piece of unstructured text is analyzed using natural language processing, each concept in the specified environment is given a score based on the way sentiment words relate to the concept and its associated score. This allows movement to a more sophisticated understanding of sentiment, because it is now possible to adjust the sentiment value of a concept relative to modifications that may surround it. Words, for example, that intensify, relax or negate the sentiment expressed by the concept can affect its score. Alternatively, texts can be given a positive and negative sentiment strength score if the goal is to determine the sentiment in a text rather than the overall polarity and strength of the text [18].

There are various other types of sentiment analysis, such as aspect-based sentiment analysis, grading sentiment analysis (positive, negative, neutral), multilingual sentiment analysis and detection of emotions [19].

This task is commonly defined as classifying a given text (usually a sentence) into one of two classes: objective or subjective. This problem can sometimes be more difficult than polarity classification. The subjectivity of words and phrases may depend on their context and an objective document may contain subjective sentences (e.g., a news article quoting people's opinions). Moreover, as mentioned by Su, results are largely dependent on the definition of subjectivity used when annotating texts. However, Pang showed that removing objective sentences from a document before classifying its polarity helped improve performance [20].

Subjective and objective identification, emerging subtasks of sentiment analysis to use syntactic, semantic features, and machine learning knowledge to identify if a sentence or document contains facts or opinions. Awareness of recognizing factual and opinions is not recent, having possibly first presented by Carbonell at Yale University in 1979 [1].

The term subjective describes the incident contains non-factual information in various forms, such as personal opinions, judgment, and predictions, also known as 'private states'. In the example down below, it reflects a private states 'We Americans'. Moreover, the target entity commented by the opinions can take several forms from tangible product to intangible topic matters stated in Liu (2010). Furthermore, three types of attitudes were observed by Liu (2010), 1) positive opinions, 2) neutral opinions, and 3) negative opinions [2].

5. Conclusion

This study underscores the profound utility of VADER for performing rapid, computationally efficient, and highly accurate sentiment analysis on massive e-commerce datasets. By integrating keyword filtering, we demonstrated a practical pathway for actionable intelligence.

Each class's collections of words or phrase indicators are

defined for to locate desirable patterns on unannotated text. For subjective expression, a different word list has been created. Lists of subjective indicators in words or phrases have been developed by multiple researchers in the linguist and natural language processing field states in Riloff et al. (2003). A dictionary of extraction rules has to be created for measuring given expressions. Over the years, in subjective detection, the features extraction progression from curating features by hand to automated features learning. At the moment, automated learning methods can further separate into supervised and unsupervised machine learning. Patterns extraction with machine learning process annotated and unannotated text have been explored extensively by academic researchers [3].

However, researchers recognized several challenges in developing fixed sets of rules for expressions respectably. Much of the challenges in rule development stems from the nature of textual information. Six challenges have been recognized by several researchers: 1) metaphorical expressions, 2) discrepancies in writings, 3) context-sensitive, 4) represented words with fewer usages, 5) time-sensitive, and 6) ever-growing volume [4].

Metaphorical expressions. The text contains metaphoric expression may impact on the performance on the extraction. Besides, metaphors take in different forms, which may have been contributed to the increase in detection [5].

Discrepancies in writings. For the text obtained from the Internet, the discrepancies in the writing style of targeted text data involve distinct writing genres and styles [6].

Time-sensitive attribute. The task is challenged by some textual data's time-sensitive attribute. If a group of researchers wants to confirm a piece of fact in the news, they need a longer time for cross-validation, than the news becomes outdated [7].

Ever-growing volume. The task is also challenged by the sheer volume of textual data. The textual data's ever-growing nature makes the task overwhelmingly difficult for the researchers to complete the task on time [8].

Previously, the research mainly focused on document level classification. However, classifying a document level suffers less accuracy, as an article may have diverse types of expressions involved. Researching evidence suggests a set of news articles that are expected to dominate by the objective expression, whereas the results show that it consisted of over 40% of subjective expression [9].

To overcome those challenges, researchers conclude that classifier efficacy depends on the precisions of patterns learner. And the learner feeds with large volumes of annotated training data outperformed those trained on less comprehensive subjective features. However, one of the main obstacles to executing this type of work is to generate a big dataset of annotated sentences manually. The manual annotation method has been less favored than automatic learning for three reasons: [10]

Variations in comprehensions. In the manual annotation task, disagreement of whether one instance is subjective or objective may occur among annotators because of languages' ambiguity [11].

All these mentioned reasons can impact on the efficiency and effectiveness of subjective and objective classification. Accordingly, two bootstrapping methods were designed to learning linguistic patterns from unannotated text data. Both methods are starting with a handful of seed words and unannotated textual data [12].

Meta-Bootstrapping by Riloff and Jones in 1999. Level One: Generate extraction patterns based on the pre-defined rules and the extracted patterns by the number of seed words each pattern holds. Level Two: Top 5 words will be marked and add to the dictionary. Repeat [13].

Basilisk (Bootstrapping Approach to Semantic Lexicon Induction using Semantic Knowledge) by Thelen and Riloff. Step One: Generate extraction patterns. Step Two: Move best patterns from Pattern Pool to Candidate Word Pool. Step Three: Top 10 words will be marked and add to the dictionary. Repeat [14].

Subjective and object classifier can enhance the several applications of natural language processing. One of the classifier's primary benefits is that it popularized the practice of data-driven decision-making processes in various industries. According to Liu, the applications of subjective and objective identification have been implemented in business, advertising, sports, and social science [15].

Online review classification: In the business industry, the classifier helps the company better understand the feedbacks on product and reasonings behind the reviews [16].

Stock price prediction: In the finance industry, the classifier aids the prediction model by process auxiliary information from social media and other textual information from the Internet. Previous studies on Japanese stock price conducted by Dong et al. indicates that model with subjective and objective module may perform better than those without this part [17].

Complex question answering. The classifier can dissect the complex questions by classing the language subject or objective and focused target. In the research Yu et al.(2003), the researcher developed a sentence and document level clustered that identity opinion pieces [18].

It refers to determining the opinions or sentiments expressed on different features or aspects of entities, e.g., of a cell phone, a digital camera, or a bank. A feature or aspect is an attribute or component of an entity, e.g., the screen of a cell phone, the service for a restaurant, or the picture quality of a camera. The advantage of feature-based sentiment analysis is the possibility to capture nuances about objects of interest. Different features can generate different sentiment responses, for example a hotel can have a convenient location, but mediocre food. This problem involves several sub-problems, e.g., identifying relevant entities, extracting their features/aspects, and determining whether an opinion expressed on each feature/aspect is positive, negative or neutral. The automatic identification of features can be performed with syntactic methods, with topic modeling, or with deep learning.

More detailed discussions about this level of sentiment analysis can be found in Liu's work [19].

Emotions and sentiments are subjective in nature. The degree of emotions/sentiments expressed in a given text at the document, sentence, or feature/aspect level-to what degree of intensity is expressed in the opinion of a document, a sentence or an entity differs on a case-to-case basis. However, predicting only the emotion and sentiment does not always convey complete information. The degree or level of emotions and sentiments often plays a crucial role in understanding the exact feeling within a single class (e.g., 'good' versus 'awesome'). Some methods leverage a stacked ensemble method for predicting intensity for emotion and sentiment by combining the outputs obtained and using deep learning models based on convolutional neural networks, long short-term memory networks and gated recurrent units [20].

5. Acknowledgement

The authors thank the Department of Computer Science and Engineering, Rungta College of Engineering and Technology, for providing the necessary resources to conduct this comprehensive review and analysis.

5. References

[1] Hutto, C., & Gilbert, E. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." *ICWSM*, 2014.

[2] Pang, B., & Lee, L. "Opinion mining and sentiment analysis." *Foundations and Trends in Information Retrieval*, 2008.

[3] Turney, P. D. "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews." *ACL*, 2002.

[4] Liu, B. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies*, 2012.

[5] Devlin, J., et al. "BERT: Pre-training of deep bidirectional transformers for language understanding." *NAACL-HLT*, 2019.

[6] Kim, Y. "Convolutional neural networks for sentence classification." *EMNLP*, 2014.

[7] Blei, D. M., et al. "Latent dirichlet allocation." *Journal of machine Learning research*, 2003.

[8] Balahur, A., et al. "Cross-lingual sentiment analysis using translation and machine learning." *Knowledge-Based Systems*, 2014.

[9] Borg, A., & Boldt, M. "Using VADER sentiment and SVM for predicting customer response sentiment." *Expert Systems with Applications*, 2020.

[10] El-Din, A., et al. "Sentiment analysis for amazon reviews using VADER and TextBlob." *AISI*, 2021.

[11] Thelwall, M., et al. "Sentiment strength detection in short informal text." *Journal of the American society for information science and technology*, 2010.

[12] Medhat, W., et al. "Sentiment analysis algorithms and applications: A survey." *Ain Shams engineering journal*, 2014.

[13] Taboada, M., et al. "Lexicon-based methods for sentiment analysis." *Computational linguistics*, 2011.

[14] Cambria, E., et al. "Sentic computing: A common-sense-based framework for concept-level sentiment analysis." *Sentiment analysis and ontology engineering*, 2016.

[15] Mohammad, S. M., & Turney, P. D. "Crowdsourcing a word-emotion association lexicon." *Computational intelligence*, 2013.

[16] Socher, R., et al. "Recursive deep models for semantic compositionality over a sentiment treebank." *EMNLP*, 2013.

[17] Hu, M., & Liu, B. "Mining and summarizing customer reviews." *KDD*, 2004.

[18] Maas, A., et al. "Learning word vectors for sentiment analysis." *ACL*, 2011.

[19] Riloff, E., et al. "Learning dictionaries for information extraction by multi-level bootstrapping." *AAAI*, 1999.

[20] Wilson, T., et al. "Recognizing contextual polarity in phrase-level sentiment analysis." *HLT-EMNLP*, 2005.